BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 March 2004 (11.03.2004)

PCT

(10) International Publication Number
WO 2004/020609 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/027401

(22) International Filing Date:
2 September 2003 (02.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/407,082 30 August 2002 (30.08.2002) US

(71) Applicant *(for all designated States except US)*: **TUFTS UNIVERSITY** [US/US]; Ballou Hall, Medford, MA 02155 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CAMILLI, Andrew** [US/US]; 5 Moose Hill Parkway, Sharon, MA 02067 (US). **HAVA, David, L.** [US/US]; 57 Quincy Way, Attleboro, MA 02703 (US).

(74) Agents: **RESNICK, David, S.** et al.; 101 Federal Street, Boston, MA 02110 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

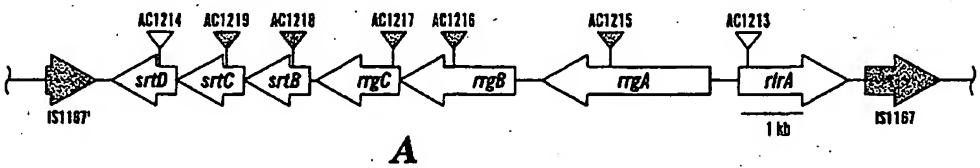
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(48) Date of publication of this corrected version:
13 May 2004

[Continued on next page]

(54) Title: STREPTOCOCCUS PNEUMONIAE ANTIGENS FOR DIAGNOSIS, TREATMENT AND PREVENTION OF ACTIVE INFECTION



A

PROTEIN	C-TERMINAL SEQUENCE
RrgA	YPRTGGIGMLPFYLIGCMMMGGVLLYTRKIIP
RrgB	IPQFGGIGTHFAVAGAAIMGIAVYAYVKNNKDEDQLA
RrgC	VPDTGEETLYILMLVAILLFGSGYYLTKKPNN

B

(57) Abstract: The present invention relates to recombinant antigenic *S. pneumoniae* polypeptides and fragments thereof. The invention also relates to methods for using these polypeptides to produce immunological responses and to confer immunological protection to disease caused by members of the genus Streptococcus, at least isolates of the *S. pneumoniae* genus. The invention further relates to nucleic acid sequences which encode antigenic *S. pneumoniae* polypeptides and to methods for detecting *S. pneumoniae* nucleic acids and polypeptides in biological samples. The invention also relates to *S. pneumoniae*-specific antibodies and methods for detecting such antibodies produced in a host animal.

WO 2004/020609 A2

(15) **Information about Correction:**
see PCT Gazette No. 20/2004 of 13 May 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Docket No. 700355-052941-PCT

Express Mail Label No. EL 948181553 US

STREPTOCOCCUS PNEUMONIAE ANTIGENS FOR DIAGNOSIS, TREATMENT AND PREVENTION OF ACTIVE INFECTION

GOVERNMENT SUPPORT

[001] This invention was supported by National Institutes of Health training grant A107422-08 and the government of the United States has certain rights thereto.

SEQUENCE LISTING

[002] The instant application contains a "lengthy" Sequence Listing which has been submitted via triplicate CD-R in lieu of a printed paper copy, and is hereby incorporated by reference in its entirety. Said CD-R, recorded on August 20, 2003, are labeled "CRF", "Copy 1" and "Copy 2", respectively, and each contains only one identical 1.38 Mb file (700941PC.APP).

FIELD OF THE INVENTION

[003] The present application is directed to *Streptococcus pneumoniae* antigens for the detection of Streptococcus, prevention of Streptococcus, and attenuation of disease caused by Streptococcus.

BACKGROUND OF THE INVENTION

[004] *Streptococcus pneumoniae* remains a major cause of morbidity and mortality in the undeveloped and developed world and resistance to common antibiotics is prevalent (2, 4, 27). *S. pneumoniae* is a component of the normal flora in the nasopharynx of approximately 50% of all adults, where it coexists with other

microflora in a nonpathogenic state. In immunocompromised people, the elderly, and young children, *S. pneumoniae* that initially colonize the nasopharynx may spread to distal sites, such as the inner ear, lower respiratory tract, or bloodstream, and cause diseases ranging from otitis media to pneumonia to meningitis (7, 18). Factors that lead to the spread from the nasopharynx to other sites of infection are not understood.

[005] Since its isolation more than 100 years ago, *Streptococcus pneumoniae* has been one of the most intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *Streptococcus pneumoniae*, however, few proteins have been identified as virulence factors involved in determining its pathogenicity.

[006] The identification of novel virulence determinants can facilitate the development of new vaccines and drug treatments, which are especially needed for organisms in which antibiotic resistance is prevalent. Antibiotic resistance is emerging rapidly in the respiratory pathogen *Streptococcus pneumoniae* (81), which is the causative agent of a number of diseases ranging in severity from normally benign otitis media to highly lethal meningitis. The major virulence factor required for all pneumococcal disease is the extracellular polysaccharide capsule, which protects colonizing or infecting bacteria from phagocytosis. In addition to the polysaccharide capsule, many surface exposed protein factors have been implicated in pneumococcal disease, however, knowledge of precise roles of many of these proteins in different *in vivo* niches is limited (21, 31).

[007] Virulence determinants of pathogens can either be essential for virulence, or not essential yet still play a role in the infection process. This distinction arises because some virulence determinants are partially or fully redundant with other determinants, and thus a mutation that prevents expression of one such determinant does not cause a noticeable attenuation of virulence because the other determinants can compensate for the missing activity/function. For example, the ability to obtain iron during infection of human tissues is a critical requirement for the virulence of most bacterial pathogens. However, many well studied pathogens have been found to employ multiple, independent systems to obtain iron from a variety of host iron-containing molecules such as heme, hemoglobin or lactoferrin. Inhibiting the

function of one pathway, for example the uptake of heme, may not cause an attenuation in virulence because the pathway for lactoferrin remains intact.

[008] Virulence determinants that can be shown to be essential in their own right are preferential targets for vaccine development and/or antimicrobial drug development. This is because the inhibition of the function of an essential determinant, for example by antibody binding or drug targeting, will reduce the potential of the pathogen to cause disease in the host. In contrast, a drug that targets and inactivates a non-essential virulence determinant will not reduce the virulence of the pathogen. Such a non-essential virulence determinant may still be targeted as a protective antigen for vaccine development, however in the face of immunological pressure, given time, the pathogen may lose the factor (by mutation) or alter the antigenicity of the non-essential determinant.

[009] Despite recent progress in indentifying genes of *Streptococcus pneumoniae,* relatively few virulence factors have been identified. Gene identification efforts include sequencing the *S. pneumoniae* serotype 4 genome. For example, U.S. 6,159,469 and U.S. 6,573,082 disclose certain *S. pneumoniae* nucleic acids potentially useful as antigens and vaccines. However, the basis for targeting these nucleic acids is merely the presence of certain sequence motifs; no data related to the biological function of the encoded proteins is provided. These nucleic acids were apparently selected as encoding proteins with signal sequences (secreted outside the cell or anchored in the membrane), lipoprotein motifs (anchored to the cell membrane), or LPXTG motifs (anchored to the cell wall outside the cell membrane).

[0010] U.S. published application 20030091577A1 discloses potentially protective protein antigens which contain signal sequences and/or LPXTG motifs. These proteins are predicted to be exported outside the bacteria and anchored onto the surface, respectively. While this patent does test a small set of peptides for protective efficacy in a mouse immunization and challenge model, only a marginal protective effect is seen for only three of the peptides tested. Thus, it is likely that the majority of these nucleic acids will fail to encode useful protective antigens if tested in the same model system.

[0011] More detailed analysis has been performed on several other proposed virulence factors. For example, WO 03/051916 A2 discloses NAD+ synthetase and paralogs as an essential protein for targeting antimicrobial drugs. WO 03/054007

discloses detailed deletion and fusion analysis to identify the most antigenic parts of two protective peptides, BVH-3 and BVH-11.

[0012] Thus, despite recent advances in sequencing by efforts, there remains a need to identify additional *S. pneumoniae* virulence factors. Signature-tagged mutagenesis (STM) represents one of several recently developed techniques for the identification of genes essential for infection (82). Several studies have identified *S. pneumoniae* virulence factors that are essential to the survival of the bacterium in different host environments by Signature-tagged mutagenesis (STM) using murine models of infection (10, 13, 23). A subset of these factors has been shown to be specific to certain host environments (10), and therefore these genes code for proteins that have tissue specific roles during infection and colonization. Among these are a number of putative transcriptional regulators, which may regulate tissue specific virulence factors in response to different host environments. However, the STM studies which have been reported are limited in terms of both the serotypes which have been screened, and the number of mutants which have been isolated. For example, two STM screens in *S. pneumoniae* have been reported, one in a serotype 19 and one in a serotype 3 strain, and have identified some virulence factors while screening only a limited number of mutants (66, 76).

[0013] Accordingly, there is a need for identification of additional *Streptococcus pneumoniae* virulence factors, from a variety of serotypes, including the isolation of the genes encoding the virulence factors, the proteins they encode, and the development of therapeutic uses of such virulence factors.

SUMMARY OF THE INVENTION

[0014] The present invention provides isolated nucleic acid molecules, having the nucleic acid sequences shown in Table 6 as SEQ ID NOs: 1-7, 10-13, 15, 17-36, 38-45, 47-49, 51, 55-56, 58-72, 74, 76-78, 80-82, 84, 86-88, 90-94, 96-97, 99-105, 107-110, 112-114, 116-122, 124-126, 128, 131-134, 136-137, 140-165, 167-170, 173-184, 187-191, 193-217, 219-222, 224-227, 229-230, and 232-237. The present invention also provides isolated polynucleotides encoding the *S. pneumoniae* polypeptides described in Table 6, and having the amino acid sequences shown in Table 6 as SEQ ID NOs.: 238-244, 247-250, 252, 254-273, 275-282, 284-286, 288, 292-293, 295-309, 311, 313-315, 317-319, 321, 323-325, 327-331, 333-334, 336-342,

344-347, 349-351, 353-359, 361-363, 365, 368-371, 373-374, 377-402, 404-407, 410-421, 424-428, 430-454, 456-459, 461-464, 466-467, and 469-474. Thus, one aspect of the invention provides isolated nucleic acid molecules comprising polynucleotides having a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding any of the amino acid sequences of the polypeptides shown as claimed in the present invention in Table 6; and (b) a nucleotide sequence complementary to any of the nucleotide sequences in (a).

[0015] Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a) or (b) above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a) or (b) above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. Additional nucleic acid embodiments of the invention relate to isolated nucleic acid molecules comprising polynucleotides which encode the amino acid sequences of epitope-bearing portions of an *S. pneumoniae* polypeptide having an amino acid sequence in (a) above.

[0016] The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using these vectors for the production of *S. pneumoniae* polypeptides or peptides by recombinant techniques.

The invention further provides isolated *S. pneumoniae* polypeptides having an amino acid sequence selected from the group consisting of an amino acid sequence of any of the polypeptides as claimed in the present invention, as shown in as claimed in the present invention in Table 6.

[0017] The polypeptides of the present invention also include polypeptides having an amino acid sequence with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to those described in as claimed in the present invention in Table 6, as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%

identical to those above; as well as isolated nucleic acid molecules encoding such polypeptides.

[0018] The present invention further provides a vaccine, preferably a multi-component vaccine comprising one or more of the *S. pneumoniae* polynucleotides or polypeptides described as claimed in the present invention in Table 6, or fragments thereof, together with a pharmaceutically acceptable diluent, carrier, or excipient, wherein the *S. pneumoniae* polypeptide(s) are present in an amount effective to elicit an immune response to members of the Streptococcus genus in an animal. The *S. pneumoniae* polypeptides of the present invention may further be combined with one or more immunogens of one or more other streptococcal or non-streptococcal organisms to produce a multi-component vaccine intended to elicit an immunological response against members of the Streptococcus genus and, optionally, one or more non-streptococcal organisms.

[0019] The vaccines of the present invention can be administered in a DNA form, e.g., "naked" DNA, wherein the DNA encodes one or more streptococcal polypeptides and, optionally, one or more polypeptides of a non-streptococcal organism. The DNA encoding one or more polypeptides may be constructed such that these polypeptides are expressed fusion proteins.

[0020] The vaccines of the present invention may also be administered as a component of a genetically engineered organism. Thus, a genetically engineered organism which expresses one or more *S. pneumoniae* polypeptides may be administered to an animal. For example, such a genetically engineered organism may contain one or more *S. pneumoniae* polypeptides of the present invention intracellularly, on its cell surface, or in its periplasmic space. Further, such a genetically engineered organism may secrete one or more *S. pneumoniae* polypeptides.

The vaccines of the present invention may be co-administered to an animal with an immune system modulator (e.g., CD86 and GM-CSF).

[0021] The invention also provides a method of inducing an immunological response in an animal to one or more members of the Streptococcus genus, preferably one or more isolates of the *S. pneumoniae* genus, comprising administering to the animal a vaccine as described above.

The invention further provides a method of inducing a protective immune response in

an animal, sufficient to prevent or attenuate an infection by members of the Streptococcus genus, preferrably at least *S. pneumoniae*, comprising administering to the animal a composition comprising one or more of the polynucleotides or polypeptides described as claimed in the present invention in Table 6, or fragments thereof. Further, these polypeptides, or fragments thereof, may be conjugated to another immunogen and/or administered in admixture with an adjuvant.

[0022] The invention further relates to antibodies elicited in an animal by the administration of one or more *S. pneumoniae* polypeptides of the present invention and to methods for producing such antibodies.

[0023] The invention also provides diagnostic methods for detecting the expression of genes of members of the Streptococcus genus in an animal. One such method involves assaying for the expression of a gene encoding *S. pneumoniae* peptides in a sample from an animal. This expression may be assayed either directly (e.g., by assaying polypeptide levels using antibodies elicited in response to amino acid sequences described as claimed in the present invention in Table 6) or indirectly (e.g., by assaying for antibodies having specificity for amino acid sequences described as claimed in the present invention in Table 6). An example of such a method involves the use of the polymerase chain reaction (PCR) to amplify and detect Streptococcus nucleic acid sequences.

[0024] The present invention also relates to nucleic acid probes having all or part of a nucleotide sequence described as claimed in the present invention in Table 6 (i.e. SEQ ID NOs: 1-7, 10-13, 15, 17-36, 38-45, 47-49, 51, 55-56, 58-72, 74, 76-78, 80-82, 84, 86-88, 90-94, 96-97, 99-105, 107-110, 112-114, 116-122, 124-126, 128, 131-134, 136-137, 140-165, 167-170, 173-184, 187-191, 193-217, 219-222, 224-227, 229-230, and 232-237) which are capable of hybridizing under stringent conditions to Streptococcus nucleic acids. The invention further relates to a method of detecting one or more Streptococcus nucleic acids in a biological sample obtained from an animal, said one or more nucleic acids encoding Streptococcus polypeptides, comprising: (a) contacting the sample with one or more of the above-described nucleic acid probes, under conditions such that hybridization occurs, and (b) detecting hybridization of said one or more probes to the Streptococcus nucleic acid present in the biological sample.

[0025] The invention also includes immunoassays, including an immunoassay for detecting Streptococcus, preferably at least isolates of the *S. pneumoniae* genus, comprising incubation of a sample (which is suspected of being infected with Streptococcus) with a probe antibody directed against an antigen/epitope of *S. pneumoniae*, to be detected under conditions allowing the formation of an antigen-antibody complex; and detecting the antigen-antibody complex which contains the probe antibody. An immunoassay for the detection of antibodies which are directed against a Streptococcus antigen comprising the incubation of a sample (containing antibodies from a mammal suspected of being infected with Streptococcus) with a probe polypeptide including an epitope of *S. pneumoniae* , under conditions that allow the formation of antigen-antibody complexes which contain the probe epitope containing antigen.

[0026] Some aspects of the invention pertaining to kits are those for: investigating samples for the presence of polynucleotides derived from Streptococcus which comprise a polynucleotide probe including a nucleotide sequence selected from the sequences shown as claimed in the present invention in Table 6 or a fragment thereof of approximately 15 or more nucleotides, in an appropriate container; analyzing the samples for the presence of antibodies directed against a Streptococcus antigen made up of a polypeptide which contains a *S. pneumoniae* epitope present in the polypeptide, in a suitable container; and analyzing samples for the presence of Streptococcus antigens made up of an anti-*S. pneumoniae* antibody, in a suitable container.

[0027] In one preferred embodiment, the present invention provides a hybridoma cell secreting a human monoclonal antibody which specifically binds to a polypeptide at least 70% identical to a sequence selected from the group consisting of an amino acid sequence of any of the polypeptides, or fragments thereof, described as claimed in the present invention in Table 6. The present invention also provides a method for generating such hybridoma cells.

[0028] In another preferred embodiment, a pharmaceutical composition is provided for reducing the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy and/or for treating *Streptococcus pneumoniae* infections comprising the human monoclonal antibody secreted by the hybridoma cell. Preferably, administration of this pharmaceutical composition can be

used for the treatment of *Streptococcus pneumoniae* infections, to reduce the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy, being an anti viral agent.

[0029] In another preferred embodiment, the human monoclonal antibody secreted by the hybridoma cell can be used for the diagnosis of *Streptococcus pneumoniae* infections in a body fluid sample.

[0030] In another preferred embodiment, the present invention provides for the use of the *Streptococcus pneumoniae* nucleic acids and polypeptides for the development of novel anti-microbial agents, and the use of such agents in the treatment and prophylaxis of *Streptococcus pneumoniae* infection.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] Figures 1A-B describe the *S. pneumoniae rlrA* locus. Figure 1A shows a schematic representation of the *S. pneumoniae rlrA* locus. *rlrA* is divergently transcribed from at least six different genes indicated by black arrows, and the entire locus is flanked by two IS1167 elements. The left element contains a frameshift mutation and is therefore predicted to be inactive. The sites of *magellan2* insertions identified by STM in *rlrA* and *srtD* is shown as open triangles and the *magellan5* insertions generated by *in vitro* transposition and used in additional animal experiments are show as black triangles. In Figure 1B, the predicted C-terminal sorting signals of RrgA, RrgB, and RrgC are listed (SEQ ID NOs: 527, 528, and 529, respectively). The LPXTG motif (SEQ ID NO: 530) of each of the proteins is conserved, with the exception of the first amino acid. All three proteins have a stretch of hydrophobic residues (underlined) and a charged tail, characteristic of proteins that are anchored to the cell wall by sortases.

[0032] Figures 2A-B show analysis of *rlrA* locus mutants in animal models. Figure 2A shows analysis of *rlrA* locus mutants in animal models of lung infection, and Figure 2B shows analysis of *rlrA* locus mutants in animal models nasopharyngeal carriage and bacteremia. The *in vivo* competitive index (CI) was calculated as described in the text; each circle represents the CI for a single mouse in each set of competitions. A CI of less than one indicates a virulence defect. Open circles indicate that no mutant bacteria were recovered from that animal and therefore 1 was

substituted in the numerator when calculating the CI. The geometric mean of the CIs for all mice in a set of competitions is shown as a solid line and statistically significant data are indicated with a symbol (* $p<0.05$, # $p<0.07$). The *in vitro* competition results for each of the tested strains are as follows: *rrgA* – 1.06, *rrgB* – 0.50, *rrgC* – 0.75, *srtB* – 0.94, *srtC* – 0.69, and *srtD* – 0.93.

[0033] Figure 3 shows a phylogenetic tree of select sortase homologues. Protein sequences of sortase homologues were aligned and a phylogenetic tree was constructed based on neighbor joining analysis. The bacterial species and, when available, the protein name are given. Bootstrap values from 100 replications are indicated at each branchpoint.

[0034] Figure 4 depicts the *rlrA* pathogenicity islet. The 12 kb locus includes a positive regulator, three surface proteins, and three sortase homologues. The four genes that are required for virulence in one or more animal models are shown in white (10).

[0035] Figures 5A-C show ribonuclease protection assays (RPA) performed to analyze the steady-state mRNA levels of each gene in the *rlrA* pathogenicity islet in both wild-type (AC353) and *rlrA* mutant strain (AC1213) backgrounds. In Figure 5A, riboprobes to each gene in the islet, as well as, to *rpoB* were generated and hybridized to 10 μg of total *S. pneumoniae* RNA from either the wild-type or mutant strain. In Figure 5B, riboprobes to *srtA* and *rpoB* were hybridized to the same samples in Figure 5A. In Figure 5C, a riboprobe that differentially recognizes the two *rlrA* transcripts in AC1278 was used to determine if RlrA is autoregulatory. The larger fragment in each lane represents the mRNA from the native *rlrA* pathogenicity islet promoter. Lanes marked with (+) are RNA samples that were harvested from cells grown in the presence of maltose.

[0036] Figures 6A-B show primer extension analysis. In Figure 6A, transcriptional start sites of promoters upstream of *rlrA* (SEQ ID NOS 531, 551 and 552, respectively, in order of appearance), *rrgA* (SEQ ID NOS 532, 553 and 554, respectively, in order of appearance), *rrgB* (SEQ ID NOS 533, 555 and 556, respectively, in order of appearance), and *srtB* (srtD) (SEQ ID NOS 534, 557 and 558, respectively, in order of appearance) were mapped by primer extension analysis (SEQ

ID NOs: 531, 532, 533, and 534. The arrow indicates primer extension products. Figure 6B is a graphical depiction of the four *rlrA* pathogenicity islet promoters. A rightward arrow indicates the +1 start site. When present, -10 and –35 σ^{70} consensus sequences and predicted Shine-Dalgarno sequences are underlined and bold.

[0037] Figures 7A-B show Northern blot analysis of *rlrA* pathogenicity islet mRNAs. Riboprobes to selected genes were synthesized and used to hybridize to total RNA recovered from AC1278 (Lane 1) or AC1213 (Lane 2) grown under maltose inducing conditions. In Figure 7A, Northern blots were probed with *rrgB* and *rrgC* riboprobes. In Figure 7B, Northern blots were probed with *srtB*, *srtC*, and *srtD* riboprobes.

[0038] Figures 8A-C show gel shift analysis using RlrA-His_6 (His tag shown in SEQ ID NO: 550). In Figure 8A, the four ^{32}P labeled probes that span the *rrgA–rlrA* intergenic region and were used in gel-shift analyses are depicted. The sizes of the PCR fragments were: AP1 – 522 bp, AP 3 – 250 bp, AP4 – 139 bp, AP5 – 163 bp, and AP7 – 290 bp. Figure 8B gel shift analysis of AP4 and AP5. ^{32}P labeled probes were incubated with increasing concentrations of RlrA- His_6 (His tag shown in SEQ ID NO: 550). The protein concentration used in each lane was: lanes 1 and 8 – 0, lanes 2 and 9 – 0.25 nM, lanes 3 and 10 – 1 nM, lanes 4 and 11 – 4 nM, lanes 5 and 12 – 16.4 nM, lanes 6 and 13 – 33 nM, and lanes 7 and 14 – 66 nM. An arrow indicates shifted species. Figure 8C shows supershift of RlrA-His_6 (His tag shown in SEQ ID NO: 550) complexes by the addition of anti-His_6 (His tag shown in SEQ ID NO: 550) antibody to the binding reaction. The concentration of protein used in each lane was: lane 1 and 4 – no protein, lane 2 and 5 – 16.4 nM RlrA-His_6 (His tag shown in SEQ ID NO: 550), and lane 3 and 6 – 16.4 nM RlrA-His_6 (His tag shown in SEQ ID NO: 550), 0.5 μg anti-His_6 (His tag shown in SEQ ID NO: 550) antibody.

[0039] Figures 9 A-B show analysis of the *rrgA–rlrA* promoter regions. Figures 9A shows DNaseI footprinting analysis of the *rrgA–rlrA* promoter regions. The ^{32}P labeled AP7 probe was incubated with increasing amounts of RlrA-His_6 (His tag shown in SEQ ID NO: 550) and subsequently treated with DNaseI. Protein concentration used in each lane was: lane 1 – 0, lane 2 – 0.5nM, lane 3 – 2.05nM, lane 4 – 8.2nM, and lane 5 – 32.8nM. DNaseI units used were: lane 1 and 2 – 0.5U,

lane 3 – 1U, lane 4 and 5 – 2U. Brackets indicate areas protected by RlrA-His_6 (His tag shown in SEQ ID NO: 550). Figure 9B depicts the *rlrA* and *rrgA* promoter regions. The oligonucleotide fragments are shown in SEQ ID NOS 535, 559 and 560, respectively, in order of appearance. RlrA binding sites are indicated in bold and the consensus binding site is underlined.

DETAILED DESCRIPTION

[0040] The present invention relates to recombinant antigenic *S. pneumoniae* polypeptides and fragments thereof. The invention also relates to methods for using these polypeptides to produce immunological responses and to confer immunological protection to disease caused by members of the genus Streptococcus, at least isolates of the *S. pneumoniae* genus. The invention further relates to nucleic acid sequences which encode antigenic *S. pneumoniae* polypeptides and to methods for detecting *S. pneumoniae* nucleic acids and polypeptides in biological samples. The invention also relates to *S. pneumoniae*-specific antibodies and methods for detecting such antibodies produced in a host animal.

[0041] The present invention takes advantage of signature tagged mutagenesis to identify novel virulence determinants of *S. pneumoniae.* Such essential proteins are excellent candidates for vaccine development and/or antimicrobial drug development, because these proteins are required for the survival and growth of the pathogen during infection. Virulence determinants of pathogens can either be essential for virulence, or not essential yet still play a role in the infection process. Virulence determinants that can be shown to be essential in their own right are preferential targets for vaccine development and/or antimicrobial drug development. This is because the inhibition of the function of an essential determinant, for example by antibody binding or drug targeting, will reduce the potential of the pathogen to cause disease in the host. In contrast, a drug that targets and inactivates a non-essential virulence determinant will not reduce the virulence of the pathogen. Such a non-essential virulence determinant may still be targeted as a protective antigen for vaccine development, however in the face of immunological pressure, given time, the pathogen may lose the factor (by mutation) or alter the antigenicity of the non-essential determinant.

[0042] The present invention provides isolated nucleic acid molecules which were identified as genes essential for lung infection by *S. pneumoniae* in a mouse

model, as described below in Example 1. Thus, the present invention provides isolated nucleic acid molecules comprising polynucleotides encoding the *S. pneumoniae* polypeptides described as claimed in the present invention in Table 6 which were determined by signature-tagged mutagenesis (STM). Table 6, below, provides information describing 237 open reading frames (ORFs) which encode potentially antigenic polypeptides of *S. pneumoniae* of the present invention. The table lists the ORF identifier assigned by the TIGR4 sequencing group, which consists of the letters SP, which denote *S. pneumoniae*, followed immediately by a four digit numeric code, which was used to arbitrarily number the *S. pneumoniae* genes. The table further correlates each ORF identifier with a sequence identification number (SEQ ID NOs:1 – 237). Thus, each ORF of the present invention is described in Table 6 first by its TIGR designation, then by the SEQ ID NO assigned to the corresponding nucleic acid sequence, SEQ ID NOs: 1 – 237, and then by the SEQ ID NO assigned to the corresponding amino acid sequence, SEQ ID NOs: 238 – 474. The actual nucleotide or amino acid sequence of each ORF identifier is also shown in the attached Sequence Disclosure under the corresponding SEQ ID NO.

[0043] Thus, for example, the designation "SP0023" refers to both the nucleotide and amino acid sequences of S. pneumoniae polypeptide number numbered 23 by the TIGR4 sequencing group. Further, "SP0023" correlates with the nucleotide sequence shown as SEQ ID NO: 1 and with the amino acid sequence shown as SEQ ID NO: 238 as is described in Table 6.

[0044] Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc.), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of DNA sequences determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a

determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

[0045] Unless otherwise indicated, each "nucleotide sequence" set forth herein is presented as a sequence of deoxyribonucleotides (abbreviated A, G , C and T). However, by "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U). For instance, reference to an RNA molecule having a sequence described in Table 6 set forth using deoxyribonucleotide abbreviations is intended to indicate an RNA molecule having a sequence in which each deoxyribonucleotide A, G or C described in Table 6 has been replaced by the corresponding ribonucleotide A, G or C, and each deoxyribonucleotide T has been replaced by a ribonucleotide U.

[0046] Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

[0047] By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0048] Isolated nucleic acid molecules of the present invention include DNA molecules comprising a nucleotide sequence described as claimed in the present invention in Table 6 and shown as SEQ ID NOs: 1 – 237; DNA molecules comprising the coding sequences for the polypeptides described as claimed in the present invention in Table 6 and shown as SEQ ID NOs:238 – 474; and DNA molecules which comprise sequences substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the *S. pneumoniae* polypeptides described as claimed in the present invention in Table 6. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate such degenerate variants.

[0049] The invention also provides nucleic acid molecules having sequences complementary to any one of those described as claimed in the present invention in Table 6. Such isolated molecules, particularly DNA molecules, are useful as probes for detecting expression of Streptococcal genes, for instance, by Northern blot analysis or the polymerase chain reaction (PCR).

[0050] The present invention is further directed to fragments of the isolated nucleic acid molecules described herein. By a fragment of an isolated nucleic acid molecule having a nucleotide sequence described in Table 6, is intended fragments at least about 15 nt, and more preferably at least about 17 nt, still more preferably at least about 20 nt, and even more preferably, at least about 25 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50-100 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of a nucleotide sequence described in Table 6. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases of a nucleotide sequence as described in Table 6. Since the nucleotide sequences identified in Table 6 are provided as SEQ ID NOs:1 – 237, generating such DNA fragments would be routine to the skilled artisan. For example, such fragments could be generated synthetically.

[0051] Preferred nucleic acid fragments of the present invention also include nucleic acid molecules comprising nucleotide sequences encoding epitope-bearing portions of the *S. pneumoniae* polypeptides identified as claimed in the present invention in Table 6. Such nucleic acid fragments of the present invention include, for example, nucleotide sequences encoding polypeptide fragments comprising from

about the amino terminal residue to about the carboxy terminal residue of each fragment shown in Table 2. The above referred to polypeptide fragments are antigenic regions of the *S. pneumoniae* polypeptides identified as claimed in the present invention in Table 6.

[0052] In another aspect, the invention provides isolated nucleic acid molecules comprising polynucleotides which hybridize under stringent hybridization conditions to a portion of a polynucleotide in a nucleic acid molecule of the invention described above, for instance, a nucleic acid sequence identified as claimed in the present invention in Table 6. By "stringent hybridization conditions" is intended overnight incubation at 42° C. in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65° C.

[0053] By polynucleotides which hybridize to a "portion" of a polynucleotide is intended polynucleotides (either DNA or RNA) which hybridize to at least about 15 nucleotides (nt), and more preferably at least about 17 nt, still more preferably at least about 20 nt, and even more preferably about 25-70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

[0054] Of course, polynucleotides hybridizing to a larger portion of the reference polynucleotide, for instance, a portion 50-100 nt in length, or even to the entire length of the reference polynucleotide, are also useful as probes according to the present invention, as are polynucleotides corresponding to most, if not all, of a nucleotide sequence as identified in Table 6. By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., a nucleotide sequences as described in Table 6). As noted above, such portions are useful diagnostically either as probes according to conventional DNA hybridization techniques or as primers for amplification of a target sequence by PCR, as described in the literature (for instance, in *Molecular Cloning, A Laboratory Manual,* 2nd. edition, Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), the entire disclosure of which is hereby incorporated herein by reference).

[0055] Since nucleic acid sequences encoding the *S. pneumoniae* polypeptides of the present invention are identified as claimed in the present invention in Table 6 and provided as SEQ ID NOs: 1-7, 10-13, 15, 17-36, 38-45, 47-49, 51, 55-56, 58-72, 74, 76-78, 80-82, 84, 86-88, 90-94, 96-97, 99-105, 107-110, 112-114, 116-122, 124-126, 128, 131-134, 136-137, 140-165, 167-170, 173-184, 187-191, 193-217, 219-222, 224-227, 229-230, and 232-237, generating polynucleotides which hybridize to portions of these sequences would be routine to the skilled artisan. For example, the hybridizing polynucleotides of the present invention could be generated synthetically according to known techniques.

[0056] As indicated, nucleic acid molecules of the present invention which encode *S. pneumoniae* polypeptides of the present invention may include, but are not limited to those encoding the amino acid sequences of the polypeptides by themselves; and additional coding sequences which code for additional amino acids, such as those which provide additional functionalities. Thus, the sequences encoding these polypeptides may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described by Gentz and colleagues (*Proc. Natl. Acad. Sci. USA* 86:821-824 (1989)), for instance, hexa-histidine provides for convenient purification of the resulting fusion protein.

[0057] Thus, the present invention also includes genetic fusions wherein the *S. pneumoniae* nucleic acid sequences coding sequences identified as claimed in the present invention in Table 6 are linked to additional nucleic acid sequences to produce fusion proteins. These fusion proteins may include epitopes of streptococcal or non-streptococcal origin designed to produce proteins having enhanced immunogenicity. Further, the fusion proteins of the present invention may contain antigenic determinants known to provide helper T-cell stimulation, peptides encoding sites for post-translational modifications which enhance immunogenicity (e.g., acylation), peptides which facilitate purification (e.g., histidine "tag"), or amino acid sequences which target the fusion protein to a desired location (e.g., a heterologous leader sequence).

[0058] In all cases of bacterial expression, an N-terminal methionine residue is added. In many cases, however, the N-terminal methionine residues is cleaved off post-translationally. Thus, the invention includes polypeptides shown in as claimed in the present invention in Table 6, with and without an N-termainal methionine.

[0059] The present invention thus includes nucleic acid molecules and sequences which encode fusion proteins comprising one or more *S. pneumoniae* polypeptides of the present invention fused to an amino acid sequence which allows for post-translational modification to enhance immunogenicity. This post-translational modification may occur either in vitro or when the fusion protein is expressed in vivo in a host cell. An example of such a modification is the introduction of an amino acid sequence which results in the attachment of a lipid moiety.

[0060] Thus, as indicated above, the present invention includes genetic fusions wherein a *S. pneumoniae* claimed nucleic acid sequence identified in Table 6 is linked to a nucleotide sequence encoding another amino acid sequence. These other amino acid sequences may be of streptococcal origin (e.g., another claimed sequence selected from Table 6) or non-streptococcal origin.

[0061] The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the *S. pneumoniae* polypeptides described as claimed in the present invention in Table 6. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (*Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985)). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

[0062] Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. These variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the *S. pneumoniae* polypeptides disclosed herein or portions thereof. Silent substitution are most likely to be made in non-epitopic regions. Guidance regarding those regions containing epitopes is provided herein, for

example, in Table 2. Also especially preferred in this regard are conservative substitutions.

[0063] Further embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to: (a) a nucleotide sequence encoding any of the amino acid sequences of the polypeptides identified as claimed in the present invention in Table 6; and (b) a nucleotide sequence complementary to any of the nucleotide sequences in (a) above.

[0064] By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a *S. pneumoniae* polypeptide described as claimed in the present invention in Table 6, is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the subject *S. pneumoniae* polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

[0065] Certain nucleotides within some of the nucleic acid sequences shown in Table 6 were ambiguous upon sequencing. Completely unknown sequences are shown as an "N". Other unresolved nucleotides are known to be either a purine, shown as "R", or a pyrimidine, shown as "Y". Accordingly, when determining identity between two nucleotide sequences, identity is met where any nucleotide, including an "R", "Y" or "N", is found in a test sequence and at the corresponding position in the referece sequence (from Table 6). Likewise, an A, G or "R" in a test sequence is identical to an "R" in the reference sequence; and a T, C or "Y" in a test sequence is identical to a "Y" in the reference sequence.

[0066] As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, a nucleotide sequence described in Table 6 can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711). Bestfit uses the local homology algorithm of Smith and Waterman (*Advances in Applied Mathematics* 2:482-489 (1981)), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

[0067] The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequences described as claimed in the present invention in Table 6. One of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention include, inter alia, (1) isolating Streptococcal genes or allelic variants thereof from either a genomic or cDNA library and (2) Northern Blot or PCR analysis for detecting Streptococcal mRNA expression.

[0068] Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to a nucleic acid sequence identified as claimed in the present invention in Table 6 will encode the same polypeptide. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay.

[0069] It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode proteins having antigenic epitopes of the *S. pneumoniae* polypeptides of the present invention. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect the antigenicity of a

polypeptide (e.g., replacement of an amino acid in a region which is not believed to form an antigenic epitope). For example, since antigenic epitopes have been identified which contain as few as six amino acids (see Harlow, et al., Antibodies: A Laboratory Manual, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1988), page 76), in instances where a polypeptide has multiple antigenic epitopes the alteration of several amino acid residues would often not be expected to eliminate all of the antigenic epitopes of that polypeptide. This is especially so when the alterations are in regions believed to not constitute antigenic epitopes.

Vectors and Host Cells

[0070] The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of *S. pneumoniae* polypeptides or fragments thereof by recombinant techniques.

[0071] Recombinant constructs may be introduced into host cells using well known techniques such as infection, transduction, transfection, transvection, electroporation and transformation. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

[0072] The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

[0073] Preferred are vectors comprising cis-acting control regions to the polynucleotide of interest. Appropriate trans-acting factors may be supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression, which may be inducible and/or cell type-specific. Particularly preferred among such vectors are those inducible by environmental factors that are easy to manipulate, such as temperature and nutrient additives.

[0074] Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors, e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episomes, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as cosmids and phagemids.

[0075] The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will preferably include a translation initiating site at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

[0076] As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracycline or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, Streptomyces and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

[0077] Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A available from Stratagene; pET series of vectors available from Novagen; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

[0078] Among known bacterial promoters suitable for use in the present invention include the *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR and PL promoters and the trp-promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I, promoter.

[0079] Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals (for example, Davis, et al., *Basic Methods In Molecular Biology* (1986)).

[0080] Transcription of DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

[0081] For secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

[0082] The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or

excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize proteins. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262).

[0083] On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as, hIL5-receptor has been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. See Bennett, D. et al., *J. Molec. Recogn.* 8:52-58 (1995) and Johanson, K. et al., *J. Biol. Chem.* 270 (16): 9459-9471 (1995).

[0084] The *S. pneumoniae* polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography and high performance liquid chromatography ("HPLC") is employed for purification. Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells.

Polypeptides and Fragments

[0085] The invention further provides isolated polypeptides having the amino acid sequences described as claimed in the present invention in Table 6, and shown as SEQ ID NOs.: 238-244, 247-250, 252, 254-273, 275-282, 284-286, 288, 292-293,

295-309, 311, 313-315, 317-319, 321, 323-325, 327-331, 333-334, 336-342, 344-347, 349-351, 353-359, 361-363, 365, 368-371, 373-374, 377-402, 404-407, 410-421, 424-428, 430-454, 456-459, 461-464, 466-467, and 469-474, and peptides or polypeptides comprising portions of the above polypeptides. The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than ten amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus.

[0086] Some amino acid sequences of the *S. pneumoniae* polypeptides described as claimed in the present invention in Table 6 can be varied without significantly effecting the antigenicity of the polypeptides. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the polypeptide which determine antigenicity. In general, it is possible to replace residues which do not form part of an antigenic epitope without significantly effecting the antigenicity of a polypeptide. Guidance for such alterations is given in Table 2 wherein epitopes for each polypeptide is delineated.

[0087] The polypeptides of the present invention are preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. Thus, a polypeptide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" is a polypeptide that has been purified, partially or substantially, from a recombinant host cell. For example, recombinantly produced versions of the *S. pneumoniae* polypeptides described in Table 6 can be substantially purified by the one-step method described by Smith and Johnson (*Gene* 67:31-40 (1988)).

[0088] The polypeptides of the present invention include: (a) an amino acid sequence of any of the polypeptides described as claimed in the present invention in Table 6; and (b) an amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a); as well as polypeptides with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to those described in (a) or (b) above, as well as polypeptides having an amino acid

sequence at least 70% identical, more preferably at least 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to those above.

[0089] By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics 2:482-489 (1981)) to find the best segment of similarity between two sequences.

[0090] By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a *S. pneumoniae* polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

[0091] The amino acid sequences shown as claimed in the present invention in Table 6 may have one or more "X" residues. "X" represents unknown. Thus, for purposes of defining identity, if any amino acid is present at the same position in a reference amino acid sequence (shown in Table 6) where an X is shown, the two sequences are identical at that position.

[0092] As a practical matter, whether any particular polypeptide is at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to, for instance, an amino acid sequence shown in Table 6 as claimed in the present invention, can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer

Group, University Research Park, 575 Science Drive, Madison, Wis. 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

[0093] As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting Streptococcal protein expression.

[0094] In another aspect, the invention provides peptides and polypeptides comprising epitope-bearing portions of the *S. pneumoniae* polypeptides of the invention. These epitopes are immunogenic or antigenic epitopes of the polypeptides of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein or polypeptide is the immunogen. These immunogenic epitopes are believed to be confined to a few loci on the molecule. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic determinant" or "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes (Geysen, et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1983)). Predicted antigenic epitopes are shown in Table 2, below.

[0095] As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein (for instance, Sutcliffe, J., et al., *Science* 219:660-666 (1983)). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the mimicked protein; longer, peptides, especially those containing proline residues, usually are effective (Sutcliffe, et al., p.

661). For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HA1 polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the rabies glycoprotein induced antibodies that precipitated the respective proteins.

[0096] Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. Thus, a high proportion of hybridomas obtained by fusion of spleen cells from donors immunized with an antigen epitope-bearing peptide generally secrete antibody reactive with the native protein (Sutcliffe, et al., p. 663). The antibodies raised by antigenic epitope-bearing peptides or polypeptides are useful to detect the mimicked protein, and antibodies to different peptides may be used for tracking the fate of various regions of a protein precursor which undergoes post-translational processing. The peptides and anti-peptide antibodies may be used in a variety of qualitative or quantitative assays for the mimicked protein, for instance in competition assays since it has been shown that even short peptides (e.g., about 9 amino acids) can bind and displace the larger peptides in immunoprecipitation assays (for instance, Wilson, et al., *Cell* 37:767-778 (1984) p. 777). The anti-peptide antibodies of the invention also are useful for purification of the mimicked protein, for instance, by adsorption chromatography using methods well known in the art.

[0097] Antigenic epitope-bearing peptides and polypeptides of the invention designed according to the above guidelines preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of a polypeptide of the invention, containing about 30 to about 50 amino acids, or any length up to and including the entire amino acid sequence of a polypeptide of the invention, also are considered epitope-bearing peptides or polypeptides of the invention and also are useful for inducing antibodies that react with the mimicked protein. Preferably, the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (i.e.,

the sequence includes relatively hydrophilic residues and highly hydrophobic sequences are preferably avoided); and sequences containing proline residues are particularly preferred.

[0098] Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Streptococcal-specific antibodies include portions of the amino acid sequences identified in Table 6 as claimed in the present invention. The polypeptide fragments disclosed in Table 6 are believed to be antigenic regions of the *S. pneumoniae* polypeptides described in Table 6. Thus the invention further includes isolated peptides and polypeptides comprising an amino acid sequence of an epitope shown in Table 6 and polynucleotides encoding said polypeptides.

[0099] The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means for making peptides or polypeptides including recombinant means using nucleic acid molecules of the invention. For instance, an epitope-bearing amino acid sequence of the present invention may be fused to a larger polypeptide which acts as a carrier during recombinant production and purification, as well as during immunization to produce anti-peptide antibodies. Epitope-bearing peptides also may be synthesized using known methods of chemical synthesis. For instance, Houghten has described a simple method for synthesis of large numbers of peptides, such as 10-20 mg of 248 different 13 residue peptides representing single amino acid variants of a segment of the HA1 polypeptide which were prepared and characterized (by ELISA-type binding studies) in less than four weeks (Houghten, R. A. *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985)). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Pat. No. 4,631,211 to Houghten and coworkers (1986). In this procedure the individual resins for the solid-phase synthesis of various peptides are contained in separate solvent-permeable packets, enabling the optimal use of the many identical repetitive steps involved in solid-phase methods. A completely manual procedure allows 500-1000 or more syntheses to be conducted simultaneously (Houghten, et al., p. 5134).

[00100] Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art (for instance, Sutcliffe, et al.; Wilson, et al.; Chow, M., et al., *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J., et al., *J. Gen. Virol.* 66:2347-2354 (1985)). Generally, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by

coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine may be coupled to carrier using a linker such as m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carrier using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats and mice are immunized with either free or carrier-coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 μg peptide or carrier protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

[00101] Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. For instance, Geysen, et al., discloses a procedure for rapid concurrent synthesis on solid supports of hundreds of peptides of sufficient purity to react in an enzyme-linked immunosorbent assay. Interaction of synthesized peptides with antibodies is then easily detected without removing them from the support. In this manner a peptide bearing an immunogenic epitope of a desired protein may be identified routinely by one of ordinary skill in the art. For instance, the immunologically important epitope in the coat protein of foot-and-mouth disease virus was located by Geysen et al. with a resolution of seven amino acids by synthesis of an overlapping set of all 208 possible hexapeptides covering the entire 213 amino acid sequence of the protein. Then, a complete replacement set of peptides in which all 20 amino acids were substituted in turn at every position within the epitope were synthesized, and the particular amino acids conferring specificity for the reaction with antibody were determined. Thus, peptide analogs of the epitope-bearing peptides of the invention can be made routinely by this method. U.S. Pat. No. 4,708,781 to Geysen (1987) further describes this method of identifying a peptide bearing an immunogenic epitope of a desired protein.

[00102] Further still, U.S. Pat. No. 5,194,392, to Geysen (1990), describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Pat. No. 4,433,092, also to Geysen (1989), describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Pat. No. 5,480,971 to Houghten, R. A. et al. (1996) discloses linear C_1-C_7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

[00103] The entire disclosure of each document cited in this section on "Polypeptides and Fragments" is hereby incorporated herein by reference.

[00104] As one of skill in the art will appreciate, the polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EPA 0,394,827; Traunecker et al., *Nature* 331:84-86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than a monomeric *S. pneumoniae* polypeptide or fragment thereof alone (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995)).

Diagnostic Assays

[00105] The present invention further relates to a method for assaying for Streptococcal infection in an animal via detecting the expression of genes encoding Streptococcal polypeptides (e.g., the polypeptides described as claimed in the present invention Table 6). This method comprises analyzing tissue or body fluid from the

animal for Streptococcus-specific antibodies or Streptococcal nucleic acids or proteins. Analysis of nucleic acid specific to Streptococcus can be done by PCR or hybridization techniques using nucleic acid sequences of the present invention as either hybridization probes or primers (cf. *Molecular Cloning: A Laboratory Manual, second edition*, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989; Eremeeva et al., *J. Clin. Microbiol.* 32:803-810 (1994) which describes differentiation among spotted fever group Rickettsiae species by analysis of restriction fragment length polymorphism of PCR-amplified DNA). Methods for detecting *B. burgdorferi* nucleic acids via PCR are described, for example, in Chen et al., *J. Clin. Microbiol.* 32:589-595 (1994).

[00106] Where diagnosis of a disease state related to infection with Streptococcus has already been made, the present invention is useful for monitoring progression or regression of the disease state whereby patients exhibiting enhanced Streptococcus gene expression will experience a worse clinical outcome relative to patients expressing these gene(s) at a lower level.

By "assaying for Streptococcal infection in an animal via detection of genes encoding Streptococcal polypeptides" is intended qualitatively or quantitatively measuring or estimating the level of one or more Streptococcus polypeptides or the level of nucleic acid encoding Streptococcus polypeptides in a first biological sample either directly (e.g., by determining or estimating absolute protein level or nucleic level) or relatively (e.g., by comparing to the Streptococcus polypeptide level or mRNA level in a second biological sample). The Streptococcus polypeptide level or nucleic acid level in the second sample used for a relative comparison may be undetectable if obtained from an animal which is not infected with Streptococcus. When monitoring the progression or regression of a disease state, the Streptococcus polypeptide level or nucleic acid level may be compared to a second sample obtained from either an animal infected with Streptococcus or the same animal from which the first sample was obtained but taken from that animal at a different time than the first. As will be appreciated in the art, once a standard Streptococcus polypeptide level or nucleic acid level which corresponds to a particular stage of a Streptococcus infection is known, it can be used repeatedly as a standard for comparison.

[00107] By "biological sample" is intended any biological sample obtained from an animal, cell line, tissue culture, or other source which contains Streptococcus

polypeptide, mRNA, or DNA. Biological samples include body fluids (such as plasma and synovial fluid) which contain Streptococcus polypeptides, and muscle, skin, and cartilage tissues. Methods for obtaining tissue biopsies and body fluids are well known in the art.

[00108] The present invention is useful for detecting diseases related to Streptococcus infections in animals. Preferred animals include monkeys, apes, cats, dogs, cows, pigs, mice, horses, rabbits and humans. Particularly preferred are humans.

[00109] Total RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, Anal. Biochem. 162:156-159 (1987). mRNA encoding Streptococcus polypeptides having sufficient homology to the nucleic acid sequences identified as claimed in the present invention in Table 6 to allow for hybridization between complementary sequences are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

[00110] Northern blot analysis can be performed as described in Harada et al., *Cell* 63:303-312 (1990). Briefly, total RNA is prepared from a biological sample as described above. For the Northern blot, the RNA is denatured in an appropriate buffer (such as glyoxal/dimethyl sulfoxide/sodium phosphate buffer), subjected to agarose gel electrophoresis, and transferred onto a nitrocellulose filter. After the RNAs have been linked to the filter by a UV linker, the filter is prehybridized in a solution containing formamide, SSC, Denhardt's solution, denatured salmon sperm, SDS, and sodium phosphate buffer. A *S. pnuemoniae* polypeptide DNA sequence shown in Table 6 as claimed in the present invention labeled according to any appropriate method (such as the ^{32}P-multiprimed DNA labeling system (Amersham)) is used as probe. After hybridization overnight, the filter is washed and exposed to x-ray film. DNA for use as probe according to the present invention is described in the sections above and will preferably at least 15 bp in length.

[00111] S1 mapping can be performed as described in Fujita et al., *Cell* 49:357-367 (1987). To prepare probe DNA for use in S1 mapping, the sense strand of an above-described *S. pnuemoniae* DNA sequence of the present invention is used as a

template to synthesize labeled antisense DNA. The antisense DNA can then be digested using an appropriate restriction endonuclease to generate further DNA probes of a desired length. Such antisense probes are useful for visualizing protected bands corresponding to the target mRNA (i.e., mRNA encoding Streptococcus polypeptides).

[00112] Preferably, levels of mRNA encoding Streptococcus polypeptides are assayed using the RT-PCR method described in Makino et al., *Technique* 2:295-301 (1990). By this method, the radioactivities of the "amplicons" in the polyacrylamide gel bands are linearly related to the initial concentration of the target mRNA. Briefly, this method involves adding total RNA isolated from a biological sample in a reaction mixture containing a RT primer and appropriate buffer. After incubating for primer annealing, the mixture can be supplemented with a RT buffer, dNTPs, DTT, RNase inhibitor and reverse transcriptase. After incubation to achieve reverse transcription of the RNA, the RT products are then subject to PCR using labeled primers. Alternatively, rather than labeling the primers, a labeled dNTP can be included in the PCR reaction mixture. PCR amplification can be performed in a DNA thermal cycler according to conventional techniques. After a suitable number of rounds to achieve amplification, the PCR reaction mixture is electrophoresed on a polyacrylamide gel. After drying the gel, the radioactivity of the appropriate bands (corresponding to the mRNA encoding the Streptococcus polypeptides)) is quantified using an imaging analyzer. RT and PCR reaction ingredients and conditions, reagent and gel concentrations, and labeling methods are well known in the art. Variations on the RT-PCR method will be apparent to the skilled artisan.

[00113] Assaying Streptococcus polypeptide levels in a biological sample can occur using any art-known method. Preferred for assaying Streptococcus polypeptide levels in a biological sample are antibody-based techniques. For example, Streptococcus polypeptide expression in tissues can be studied with classical immunohistological methods. In these, the specific recognition is provided by the primary antibody (polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunohistological staining of tissue section for pathological examination is obtained. Tissues can also be extracted, e.g., with urea and neutral detergent, for the liberation of Streptococcus polypeptides for Western-blot or dot/slot assay (Jalkanen,

M., et al., *J. Cell. Biol.* 101:976-985 (1985); Jalkanen, M., et al., *J. Cell. Biol.* 105:3087-3096 (1987)). In this technique, which is based on the use of cationic solid phases, quantitation of a Streptococcus polypeptide can be accomplished using an isolated Streptococcus polypeptide as a standard. This technique can also be applied to body fluids.

[00114] Other antibody-based methods useful for detecting Streptococcus polypeptide gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). For example, a Streptococcus polypeptide-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify a Streptococcus polypeptide. The amount of a Streptococcus polypeptide present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. Such an ELISA for detecting a tumor antigen is described in Iacobelli et al., *Breast Cancer Research and Treatment* 11:19-30 (1988). In another ELISA assay, two distinct specific monoclonal antibodies can be used to detect Streptococcus polypeptides in a body fluid. In this assay, one of the antibodies is used as the immunoabsorbent and the other as the enzyme-labeled probe.

[00115] The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting the Streptococcus polypeptide with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample.
Streptococcus polypeptide-specific antibodies for use in the present invention can be raised against an intact S. pneumoize polypeptide of the present invention or fragment thereof. These polypeptides and fragments may be administered to an animal (e.g., rabbit or mouse) either with a carrier protein (e.g., albumin) or, if long enough (e.g., at least about 25 amino acids), without a carrier.

[00116] As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and $F(ab')_2$ fragments) which are capable of specifically binding to a

Streptococcus polypeptide. Fab and F(ab')$_2$ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

[00117] The antibodies of the present invention may be prepared by any of a variety of methods. For example, the *S. pneumoniae* polypeptides identified in Table 6 as claimed in the present invention, or fragments thereof, can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of a *S. pneumoniae* polypeptide of the present invention is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of high specific activity.

[00118] In the most preferred method, the antibodies of the present invention are monoclonal antibodies. Such monoclonal antibodies can be prepared using hybridoma technology (Kohler et al., Nature 256:495 (1975); Kohler et al., *Eur. J. Immunol.* 6:511 (1976); Kohler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., In: *Monoclonal Antibodies and T-Cell Hybridomas,* Elsevier, N.Y., (1981) pp. 563-681). In general, such procedures involve immunizing an animal (preferably a mouse) with a *S. pneumoniae* polypeptide antigen of the present invention. Suitable cells can be recognized by their capacity to bind anti-Streptococcus polypeptide antibody. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56° C.), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP 20), available from the American Type Culture Collection, Rockville, Md. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the Streptococcus polypeptide antigen administered to immunized animal.

[00119] Alternatively, additional antibodies capable of binding to Streptococcus polypeptide antigens may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, Streptococcus polypeptide-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the Streptococcus polypeptide-specific antibody can be blocked by a Streptococcus polypeptide antigen. Such antibodies comprise anti-idiotypic antibodies to the Streptococcus polypeptide-specific antibody and can be used to immunize an animal to induce formation of further Streptococcus polypeptide-specific antibodies.

[00120] It will be appreciated that Fab and $F(ab')_2$ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce $F(ab')_2$ fragments). Alternatively, Streptococcus polypeptide-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

[00121] Of special interest to the present invention are antibodies to Streptococcus polypeptide antigens which are produced in humans, or are "humanized" (i.e., non-immunogenic in a human) by recombinant or other technology. Humanized antibodies may be produced, for example by replacing an immunogenic portion of an antibody with a corresponding, but non-immunogenic portion (i.e., chimeric antibodies) (Robinson, R. R. et al., International Patent Publication PCT/US86/02269; Akira, K. et al., European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison, S. L. et al., European Patent Application 173,494; Neuberger, M. S. et al., PCT Application WO 86101533; Cabilly, S. et al., European Patent Application 125,023; Better, M. et al., Science 240:1041-1043 (1988); Liu, A. Y. et al., *Proc. Natl. Acad. Sci. USA* 84:3439-3443 (1987); Liu, A. Y. et al., *J. Immunol.* 139:3521-3526 (1987); Sun, L. K. et al., *Proc. Natl. Acad. Sci. USA* 84:214-218 (1987); Nishimura, Y. et al., *Canc. Res.* 47:999-1005 (1987); Wood, C. R. et al., *Nature* 314:446-449 (1985)); Shaw et al., *J. Natl. Cancer Inst.* 80:1553-1559 (1988). General reviews of "humanized" chimeric antibodies are provided by Morrison, S. L.

(*Science,* 229:1202-1207 (1985)) and by Oi, V. T. et al., *BioTechniques* 4:214 (1986)). Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones, P. T. et al., Nature 321:552-525 (1986); Verhoeyan et al., *Science* 239:1534 (1988); Beidler, C. B. et al., *J. Immunol.* 141:4053-4060 (1988)).

[00122] Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Glucose oxidase is particularly preferred as it has good stability and its substrate (glucose) is readily available. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labeled antibody/substrate reaction. Besides enzymes, other suitable labels include radioisotopes, such as iodine (^{125}I, ^{121}I), carbon (^{14}C), sulphur (^{35}S), tritium (^{3}H), indium (^{112}In), and technetium (^{99m}Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

[00123] Further suitable labels for the Streptococcus polypeptide-specific antibodies of the present invention are provided below. Examples of suitable enzyme labels include malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, and acetylcholine esterase.

[00124] Examples of suitable radioisotopic labels include ^{3}H, ^{111}In, ^{125}I, ^{131}I, ^{32}P, ^{35}S, ^{14}C, ^{51}Cr, ^{57}To, ^{58}Co, ^{59}Fe, ^{75}Se, ^{152}Eu, ^{90}Y, ^{67}Cu, ^{217}Ci, ^{211}At, ^{212}Pb, ^{47}Sc, ^{109}Pd, etc. In is a preferred isotope where in vivo imaging is used since its avoids the problem of dehalogenation of the ^{125}I or ^{131}I-labeled monoclonal antibody by the liver. In addition, this radionucleotide has a more favorable gamma emission energy for imaging (Perkins et al., *Eur. J. Nucl. Med.* 10:296-301 (1985); Carasquillo et al., *J. Nucl. Med.* 28:281-287 (1987)). For example, ^{111}In coupled to monoclonal antibodies with 1-(P-isothiocyanatobenzyl)-DPTA has shown little uptake in non-tumorous tissues, particularly the liver, and therefore enhances specificity of tumor localization (Esteban et al., *J. Nucl. Med.* 28:861-870 (1987)).

[00125] Examples of suitable non-radioactive isotopic labels include ^{157}Gd, ^{55}Mn, ^{162}Dy, ^{52}Tr, and ^{56}Fe.

[00126] Examples of suitable fluorescent labels include an ^{152}Eu label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, and a fluorescamine label.

[00127] Examples of suitable toxin labels include diphtheria toxin, ricin, and cholera toxin.

[00128] Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, and an aequorin label.

[00129] Examples of nuclear magnetic resonance contrasting agents include heavy metal nuclei such as Gd, Mn, and iron.

[00130] Typical techniques for binding the above-described labels to antibodies are provided by Kennedy et al., *Clin. Chim. Acta* 70:1-31 (1976), and Schurs et al., *Clin. Chim. Acta* 81:1-40 (1977). Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

[00131] In a related aspect, the invention includes a diagnostic kit for use in screening serum containing antibodies specific against *S. pneumoniae* infection. Such a kit may include an isolated *S. pneumoniae* antigen comprising an epitope which is specifically immunoreactive with at least one anti-S. -pneumoniae antibody. Such a kit also includes means for detecting the binding of said antibody to the antigen. In specific embodiments, the kit may include a recombinantly produced or chemically synthesized peptide or polypeptide antigen. The peptide or polypeptide antigen may be attached to a solid support.

[00132] In a more specific embodiment, the detecting means of the above-described kit includes a solid support to which said peptide or polypeptide antigen is attached. Such a kit may also include a non-attached reporter-labelled anti-human antibody. In this embodiment, binding of the antibody to the *S. pneumoniae* antigen can be detected by binding of the reporter labelled antibody to the anti-*S. pneumoniae* antibody.

[00133] In a related aspect, the invention includes a method of detecting *S. pneumoniae* infection in a subject. This detection method includes reacting a body fluid, preferably serum, from the subject with an isolated *S. pneumoniae* antigen, and examining the antigen for the presence of bound antibody. In a specific embodiment, the method includes a polypeptide antigen attached to a solid support, and serum is reacted with the support. Subsequently, the support is reacted with a reporter-labeled anti-human antibody. The support is then examined for the presence of reporter-labeled antibody.

[00134] The solid surface reagent employed in the above assays and kits is prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, 96-well plates or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent attachment of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activated carboxyl, hydroxyl, or aldehyde group. Alternatively, streptavidin coated plates can be used in conjunction with biotinylated antigen(s).

Human Monoclonal Antibodies

[00135] One preferred embodiment of the present invention provides human monoclonal antibodies to *Streptococcus pneumoniae* antigens. Methods for producing human monoclonal antibodies in human/mouse chimeras are known in the art, for example in U.S. Patent No. 6,254,867, which is hereby incorporated by reference herein.

[00136] The present invention provides for generating hybridoma cell lines that secrete human monoclonal antibodies that specifically bind to any of the *S. pneumoniae* polypeptides, or fragments thereof, described in Table 6 as claimed in the present invention.

[00137] The antigen used for immunizing the chimeric rodent is preferably any one or more of the *S. pneumoniae* polypeptides, or fragments thereof, described in Table 6 as claimed in the present invention. The antigen for example may be prepared as a suspension adsorbed on aluminum hydroxide.

Therapeutics and Modes of Administration

[00138] The present invention also provides vaccines comprising one or more polypeptides of the present invention. Heterogeneity in the composition of a vaccine may be provided by combining *S. pneumoniae* polypeptides of the present invention. Multi-component vaccines of this type are desirable because they are likely to be more effective in eliciting protective immune responses against multiple species and strains of the Streptococcus genus than single polypeptide vaccines. Thus, as discussed in detail below, a multi-component vaccine of the present invention may contain one or more, preferably 2 to about 20, more preferably 2 to about 15, and most preferably 3 to about 8, of the *S. pneumoniae* polypeptides identified in Table 6 as claimed in the present invention, or fragments thereof.

[00139] Multi-component vaccines are known in the art to elicit antibody production to numerous immunogenic components. Decker, M. and Edwards, K., *J. Infect. Dis.* 174: S270-275 (1996). In addition, a hepatitis B, diphtheria, tetanus, pertussis tetravalent vaccine has recently been demonstrated to elicit protective levels of antibodies in human infants against all four pathogenic agents. Aristegui, J. et al., *Vaccine* 15:7-9 (1997).

[00140] The present invention thus also includes multi-component vaccines. These vaccines comprise more than one polypeptide, immunogen or antigen. An example of such a multi-component vaccine would be a vaccine comprising more than one of the *S. pneumoniae* polypeptides described in Table 6 as claimed in the present invention. A second example is a vaccine comprising one or more, for example 2 to 10, of the *S. pneumoniae* polypeptides identified in Table 6 as claimed in the present invention, and one or more, for example 2 to 10, additional polypeptides of either streptococcal or non-streptococcal origin. Thus, a multi-component vaccine which confers protective immunity to both a Streptococcal infection and infection by another pathogenic agent is also within the scope of the invention.

[00141] Further within the scope of the invention are whole cell and whole viral vaccines. Such vaccines may be produced recombinantly and involve the expression of one or more of the *S. pneumoniae* polypeptides described in Table 6 as claimed in the present invention. For example, the *S. pneumoniae* polypeptides of the present invention may be either secreted or localized intracellular, on the cell surface, -or in the periplasmic space. Further, when a recombinant virus is used, the S. pneumoniae polypeptides of the present invention may, for example, be localized in the viral

envelope, on the surface of the capsid, or internally within the capsid. Whole cells vaccines which employ cells expressing heterologous proteins are known in the art. See, e.g., Robinson, K. et al., *Nature Biotech.* 15:653-657 (1997); Sirard, J. et al., *Infect. Immun.* 65:2029-2033 (1997); Chabalgoity, J. et al., *Infect. Immun.* 65:2402-2412 (1997). These cells may be administered live or may be killed prior to administration. Chabalgoity, J. et al., for example, report the successful use in mice of a live attenuated Salmonella vaccine strain which expresses a portion of a platyhelminth fatty acid-binding protein as a fusion protein on its cells surface.

[00142] A multi-component vaccine can also be prepared using techniques known in the art by combining one or more *S. pneumoniae* polypeptides of the present invention, or fragments thereof, with additional non-streptococcal components (e.g., diphtheria toxin or tetanus toxin, and/or other compounds known to elicit an immune response). Such vaccines are useful for eliciting protective immune responses to both members of the Streptococcus genus and non-streptococcal pathogenic agents The vaccines of the present invention also include DNA vaccines. DNA vaccines are currently being developed for a number of infectious diseases. Boyer, J et al., *Nat. Med.* 3:526-532 (1997); reviewed in Spier, R., *Vaccine* 14:1285-1288 (1996). Such DNA vaccines contain a nucleotide sequence encoding one or more *S. pneumoniae* polypeptides of the present invention oriented in a manner that allows for expression of the subject polypeptide. The direct administration of plasmid DNA encoding *B. burgdorgeri* OspA has been shown to elicit protective immunity in mice against borrelial challenge. Luke, C. et al., *J. Infect. Dis.* 175:91-97 (1997).

[00143] The present invention also relates to the administration of a vaccine which is co-administered with a molecule capable of modulating immune responses. Kim, J. et al., *Nature Biotech.* 15:641-646 (1997), for example, report the enhancement of immune responses produced by DNA immunizations when DNA sequences encoding molecules which stimulate the immune response are co-administered. In a similar fashion, the vaccines of the present invention may be co-administered with either nucleic acids encoding immune modulators or the immune modulators themselves. These immune modulators include granulocyte macrophage colony stimulating factor (GM-CSF) and CD86.

[00144] The vaccines of the present invention may be used to confer resistance to streptococcal infection by either passive or active immunization. When the vaccines

of the present invention are used to confer resistance to streptococcal infection through active immunization, a vaccine of the present invention is administered to an animal to elicit a protective immune response which either prevents or attenuates a streptococcal infection. When the vaccines of the present invention are used to confer resistance to streptococcal infection through passive immunization, the vaccine is provided to a host animal (e.g., human, dog, or mouse), and the antisera elicited by this antisera is recovered and directly provided to a recipient suspected of having an infection caused by a member of the Streptococcus genus.

[00145] The ability to label antibodies, or fragments of antibodies, with toxin molecules provides an additional method for treating streptococcal infections when passive immunization is conducted. In this embodiment, antibodies, or fragments of antibodies, capable of recognizing the *S. pneumoniae* polypeptides disclosed herein, or fragments thereof, as well as other Streptococcus proteins, are labeled with toxin molecules prior to their administration to the patient. When such toxin derivatized antibodies bind to Streptococcus cells, toxin moieties will be localized to these cells and will cause their death.

[00146] The present invention thus concerns and provides a means for preventing or attenuating a streptococcal infection resulting from organisms which have antigens that are recognized and bound by antisera produced in response to the polypeptides of the present invention. As used herein, a vaccine is said to prevent or attenuate a disease if its administration to an animal results either in the total or partial attenuation (i.e., suppression) of a symptom or condition of the disease, or in the total or partial immunity of the animal to the disease.

[00147] The administration of the vaccine (or the antisera which it elicits) may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the compound(s) are provided in advance of any symptoms of streptococcal infection. The prophylactic administration of the compound(s) serves to prevent or attenuate any subsequent infection. When provided therapeutically, the compound(s) is provided upon or after the detection of symptoms which indicate that an animal may be infected with a member of the Streptococcus genus. The therapeutic administration of the compound(s) serves to attenuate any actual infection. Thus, the *S. pneumoniae* polypeptides, and fragments thereof, of the present invention may be provided either

prior to the onset of infection (so as to prevent or attenuate an anticipated infection) or after the initiation of an actual infection.

[00148] The polypeptides of the invention, whether encoding a portion of a native protein or a functional derivative thereof, may be administered in pure form or may be coupled to a macromolecular carrier. Example of such carriers are proteins and carbohydrates. Suitable proteins which may act as macromolecular carrier for enhancing the immunogenicity of the polypeptides of the present invention include keyhole limpet hemacyanin (KLH) tetanus toxoid, pertussis toxin, bovine serum albumin, and ovalbumin. Methods for coupling the polypeptides of the present invention to such macromolecular carriers are disclosed in Harlow et al., *Antibodies: A Laboratory Manual,* 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1988), the entire disclosure of which is incorporated by reference herein.

[00149] A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient animal and is otherwise suitable for administration to that animal. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

[00150] While in all instances the vaccine of the present invention is administered as a pharmacologically acceptable compound, one skilled in the art would recognize that the composition of a pharmacologically acceptable compound varies with the animal to which it is administered. For example, a vaccine intended for human use will generally not be co-administered with Freund's adjuvant. Further, the level of purity of the *S. pneumoniae* polypeptides of the present invention will normally be higher when administered to a human than when administered to a non-human animal.

[00151] As would be understood by one of ordinary skill in the art, when the vaccine of the present invention is provided to an animal, it may be in a composition which may contain salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. Adjuvants are substances that can be used to specifically augment a specific immune response. These substances generally

perform two functions: (1) they protect the antigen(s) from being rapidly catabolized after administration and (2) they nonspecifically stimulate immune responses.

[00152] Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same site of the animal being immunized. Adjuvants can be loosely divided into several groups based upon their composition. These groups include oil adjuvants (for example, Freund's complete and incomplete), mineral salts (for example, $AlK(SO_4)_2$, $AlNa(SO_4)_2$, $AlNH_4(SO_4)$, silica, kaolin, and carbon), polynucleotides (for example, poly IC and poly AU acids), and certain natural substances (for example, wax D from *Mycobacterium tuberculosis*, as well as substances found in *Corynebacterium parvum*, or *Bordetella pertussis*, and members of the genus Brucella. Other substances useful as adjuvants are the saponins such as, for example, Quil A. (Superfos A/S, Denmark). Preferred adjuvants for use in the present invention include aluminum salts, such as $AlK(SO_4)_2$, $AlNa(SO_4)_2$, and $AlNH_4(SO_4)$. Examples of materials suitable for use in vaccine compositions are provided in Remington's Phannaceutical Sciences (Osol, A, Ed, Mack Publishing Co, Easton, Pa., pp. 1324-1341 (1980), which reference is incorporated herein by reference).

[00153] The therapeutic compositions of the present invention can be administered parenterally by injection, rapid infusion, nasopharyngeal absorption (intranasopharangeally), dermoabsorption, or orally. The compositions may alternatively be administered intramuscularly, or intravenously. Compositions for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Carriers or occlusive dressings can be used to increase skin permeability and enhance antigen absorption. Liquid dosage forms for oral administration may generally comprise a liposome solution containing the liquid dosage form. Suitable forms for suspending liposomes include emulsions, suspensions, solutions, syrups, and elixirs containing inert diluents commonly used in the art, such as purified water. Besides the inert diluents, such compositions can also include adjuvants, wetting agents, emulsifying and suspending agents, or sweetening, flavoring, or perfuming agents.

[00154] Therapeutic compositions of the present invention can also be administered in encapsulated form. For example, intranasal immunization of mice against *Bordetella pertussis* infection using vaccines encapsulated in biodegradable microsphere composed of poly(DL-lactide-co-glycolide) has been shown to stimulate protective immune responses. Shahin, R. et al., *Infect. Immun.* 63:1195-1200 (1995). Similarly, orally administered encapsulated Salmonella typhimurium antigens have also been shown to elicit protective immunity in mice. Allaoui-Attarki, K. et al., *Infect. Immun.* 65:853-857 (1997). Encapsulated vaccines of the present invention can be administered by a variety of routes including those involving contacting the vaccine with mucous membranes (e.g., intranasally, intracolonicly, intraduodenally). Many different techniques exist for the timing of the immunizations when a multiple administration regimen is utilized. It is possible to use the compositions of the invention more than once to increase the levels and diversities of expression of the immunoglobulin repertoire expressed by the immunized animal. Typically, if multiple immunizations are given, they will be given one to two months apart.

[00155] According to the present invention, an "effective amount" of a therapeutic composition is one which is sufficient to achieve a desired biological effect. Generally, the dosage needed to provide an effective amount of the composition will vary depending upon such factors as the animal's or human's age, condition, sex, and extent of disease, if any, and other variables which can be adjusted by one of ordinary skill in the art.

[00156] The antigenic preparations of the invention can be administered by either single or multiple dosages of an effective amount. Effective amounts of the compositions of the invention can vary from 0.01-1,000 μg/ml per dose, more preferably 0.1-500 μg/ml per dose, and most preferably 10-300 μg/ml per dose.

[00157] The present invention also provides methods for identifying potential anti-microbial agents capable of antagonizing, inhibiting or otherwise interfering with the function of a polypeptide of SEQ ID NO:238-474. One preferred method provides for inactivating the polypeptide in *Streptococcus pneumoniae*, exposing the strain to a candidate agent, and determining whether the *Streptococcus pneumoniae* is still viable in vitro or in vivo. Another preferred method for the identification of an agent that is effective in the treatment and/or diagnosis of *Streptococcus pneumoniae* infection,

provides contacting a polypeptide of SEQ ID NO: 238-474 with a target agent, and selecting an agent that binds specifically to said nucleic acid or polypeptide.

[00158] Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced.

[00159] Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means. In addition, known pharmacological agents may be subject to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc.

[00160] Having now generally described the invention, the same will be more readily understood through reference to the following example which is provided by way of illustration, and is not intended to be limiting of the present invention, unless specified.

EXAMPLES

EXAMPLE 1: Large Scale Identification of Serotype 4 *Streptococcus pneumoniae* Virulence Factors

MATERIALS AND METHODS

Bacterial strains, plasmids, and DNA manipulations

[00161] Strains, plasmids, and primers used in this study are listed in Table 3. All *S. pneumoniae* strains used and constructed in this study are derivatives of TIGR4, a serotype 4 clinical isolate. Antibiotic concentrations used in this study were as follows: chloramphenicol (Cm) 4 $\mu g\ ml^{-1}$, streptomycin (Sm) 100 $\mu g\ ml^{-1}$, and spectinomycin (Spc) 200 $\mu g\ ml^{-1}$ for *S. pneumoniae*; Cm 10 $\mu g\ ml^{-1}$ and Spc 100 $\mu g\ ml^{-1}$ for *E. coli*. All DNA manipulations were carried out according to standard protocols (Sambrook *et al.*, 1998). Signature-tags were PCR amplified from a plasmid preparation of pUTmTn5Km2 (Hensel *et al.*, 1995) using primers P6 and P7.

PCR amplification conditions were as follows: 30 cycles of 96°C for 30 s, 94°C for 20 s, 52°C for 45 s, and 72°C for 10 s, followed by a final dwell at 72 °C for 15 minutes. PCR products were ethanol precipitated and resuspended in Bgl II buffer (New England Biolabs), and digested with Bgl II overnight at 37°C. Plasmid pEMcat was digested overnight with Bgl II. Both the linearized plasmid and the signature-tags were gel purified, and the former was dephosphorylated using shrimp alkaline phosphatase according to the manufacturer's instructions (Boehringer Mannheim). Purified signature-tags were ligated into the vector overnight with T4 DNA ligase (New England Biolabs) and the ligation mixture was introduced into *E. coli* DH5αλpir via electroporation. Transformants were selected on Luria-bertani (LB) agar plates supplemented with Cm.

[00162] Transformants that contained uniquely tagged mini-transposons on pEMcat were isolated as follows. Colony purified transformants were grown overnight in four microtiter plates, each plate comprised a pool. Next, 4 μl of each well was spotted onto Duralon nitrocellulose membranes, one pool per membrane (Stratagene). Membranes (one per pool) were transferred onto filter paper (Whatman) saturated in denaturation solution (0.5N NaOH, 1.5M NaCl) for 10 minutes (mins), 0.1% SDS for 3 min, and lastly, neutralization solution [1.0M Tris HCl (pH 7.5), 1.5M NaCl] for 3 min, at which time, DNA was cross-linked to membranes in a UV Stratalinker (Stratagene). The membranes were incubated in 3X SSC, 0.1% SDS for 1h, and cellular debris was gently removed from the membranes by rubbing with Kimwipes (Kimberly-Clarke). Probe was generated from each pool using primers P6 and P7 by dioxygenin(DIG)-dUTP labeling PCR as described by the manufacturer (Roche). Cross-reacting signature-tags were eliminated between each of the pools by successive hybridizations of probe from one pool to blots with tags from another pool. From these hybridizations, 129 strains that did not cross-hybridize were randomly assembled into two new pools, and screened for cross-hybridizing signature-tags as above. Finally, 93 strains were selected that did not cross-hybridize and that contained signature-tags were isolated that amplify well by PCR. To generate master dot blots for hybridization of input and output signature tags, the unique 40 bp signature-tag of each *magellan2* transposon was purified and spotted onto membranes as described (Merrell *et al.*, 2002). All membranes were stored at 4°C.

In vitro *transposon mutagenesis, DNA transformation, and pool construction*

[00163] Plasmid DNA was purified from *E. coli* strains harboring each of the 93 uniquely tagged *magellan2* elements using Qiagen mini plasmid preparation kit according to the manufacturer's instructions (Qiagen). *S. pneumoniae* genomic DNA was isolated from AC353 as follows: AC353 was grown in 40 ml of THY (Todd Hewitt broth, 0.5% yeast extract) supplemented with Sm and 5 μl ml^{-1} Oxyrase (Oxyrase Inc.) statically in a candle extinction jar. Cells were washed in sterile dH_20, resuspended in 200 μl of lysis buffer (0.1% deoxycholate, 0.01% SDS, 0.15M NaCl) and incubated at 37°C for 10 min. Next, 0.9 ml of SSC was added and samples were incubated an additional 10 min at 65°C. The cell lysate was phenol extracted, chloroform extracted and ethanol precipitated. Precipitated DNA was washed in 70% ethanol, and resuspended in 200 μl of 50mM Tris-HCl (pH 7.5), 5mM $CaCl_2$. 10 μl of proteinase K (10 mg ml^{-1}) and 2 μl of RNase (100 mg ml^{-1}) were added and the mixture was incubated at 37°C for 10 min. EDTA was added to 10mM to stop the reaction, and the lysate was again extracted with phenol and chloroform, and ethanol precipitated.

[00164] *In vitro magellan2* transposition reactions were carried out with purified MarC9 transposase, 500 ng of target AC353 genomic DNA and 1 μg of each of pEMcat derivative separately, essentially as described (Lampe *et al.*, 1999). Reactions were ethanol precipitated and resuspended in gap repair buffer [50mM Tris (pH 7.8), 10mM $MgCl_2$, 1mM DTT, 100nM dNTP, and 50 ng of BSA]. Repair of transposition product gaps was performed as described (Akerley *et al.*, 1998), except that *E. coli* DNA ligase (NEB) was used in place of T4 DNA ligase. Repaired transposition products were transformed into naturally competent AC353 as described (Bricker and Camilli, 1999). Of the 93 pEMcat derivatives used in the above procedure, only 63 reproducibly yielded sufficient numbers of transformants. Therefore, Cm^R colonies were picked from these 63 transformations only, and statically grown to late logarithmic phase in 200 μl of THY in 96 well microtiter plates in candle extinction jars, and subsequently frozen after the addition of glycerol to 20% (v/v). This entire procedure was repeated three times to assemble 100 pools of 63 mutant strains to be used for STM screening as described below. For the assembly of 2° pools, 1 μl of frozen cells from the appropriate well was inoculated

into 200 µl of THY in a 96 well microtiter plate and grown for 5 h to log phase as above. Glycerol was added to 20% (v/v) and the plates were stored at -75 °C.

[00165] *Magellan5* transposon insertions into the *rlrA* locus were generated identically to the *magellan2* mutagenesis, except that two different 7 kb PCR products were used as target DNA. PCR products were amplified from AC353 with primer sets TNPAB-F/REG2-R and REG2-F/PFL-R (Table 4) and purified using the Qiagen PCR purification kit according to manufacturers guidelines. *In vitro* transposition, gap repair, and natural transformation were carried out exactly as for *magellan2*.

Animal infections

[00166] In all animal infections 6 - 10 week-old female Swiss Webster mice were used (Taconic Labs). Mice were provided with continuous food and water, and housed according to the Tufts University Department of Lab Animal Medicine guidelines. Pools were prepared for infection by resuspending ~1 µl of frozen cells in 25 µl of THY, and plating 5 µl of each strain as a discrete spot on a blood agar plate [Blood Agar Base No. 2 (Difco) and 5% defibrinated sheep blood] supplemented with Cm and Sm. Following overnight growth, the entire pool was resuspended in THY and adjusted to $OD_{600} \approx 0.85$ (approximately 5×10^8 CFU ml^{-1}), the remainder of this resuspension was saved and used to asses the complexity of the input population of bacteria (see below). In the determination of colonization bottlenecks, and the 1° and 2° STM screens, 40 µl of each resuspended pool was inoculated intranasally into two lightly anesthetized mice using methoxyflurane inhalation. The infections were carried out for 44 h at which time, mice were sacrificed by CO_2 asphyxiation. Both lungs from each animal were aseptically removed, and homogenized in 5 ml of THY-glycerol (20% v/v) and stored at -75°C.

[00167] Serial dilutions of bacteria recovered from each mouse were plated on blood plates supplemented with Sm and Cm, such that a semiconfluent lawn of colonies was obtained. Bacteria were recovered with THY, genomic DNA from input and output bacteria were prepared using the DNAEasy Tissue kit according to the manufacturers tissue preparation protocol (Qiagen). Recovered genomic DNA was used as template for PCR amplification of the signature tags. DIG-dUTP was

incorporated during the PCR as described above and signature-tag master blots were probed as described (Merrell *et al.*, 2002).

Competition Experiments

[00168] Prior to competition experiments, *magellan2* insertion mutations were back-crossed into AC353 as follows: Genomic DNA was prepared from each selected mutant strain as above, and was used to transform natural competent AC353 as described (Bricker and Camilli, 1999). Mutant and wild-type (AC353) strains were grown separately on blood agar plates with appropriate antibiotics, and recovered and prepared for infection identically to the input pools above. Prior to infection, mutant bacteria and AC353 were mixed in a 1:1 ratio, and inoculated at the following doses: 1×10^7 CFU for lung infections, 5×10^5 for *i.p.* infections, and 1×10^8 CFU for nasopharyngeal inoculation. *I.p.* infections were carried out for 20 h and nasopharyngeal carriage infections for 7 days. Bacteria from systemic infections were recovered from the bloodstream by cardiac puncture. Nasopharyngeal colonized bacteria were recovered by washing the nasopharynx with 400 µl of sterile phosphate buffered saline essentially as described (Wu *et al.*, 1997). In conjunction with each *in vivo* competition, an *in vitro* competition was carried out as follows: 40 µl of each mixture was inoculated into 10 ml of THY supplemented with Sm (50 µg ml^{-1}) and Oxyrase (5 µl ml^{-1}) and grown (~9 doublings) to mid-log phase for 5 h. Following each experiment, the ratio of mutant to wild-type bacteria, for both *in vitro* and *in vivo* competitions, was determined by first plating recovered bacteria on TSA blood plates with Sm, and subsequently replica-plating colonies to plates with Sm or Sm and Cm. Competitive indices were calculated as the ratio of mutant to wild-type bacteria recovered from each animal (*in vivo* CI) or from THY broth (*in vitro* CI) adjusted by the input ratio.

Arbitrary-primed PCR, DNA sequencing, and Sequence analysis

[00169] For each of the 387 strains determined to be highly attenuated by STM screening, we attempted to amplify one *magellan2*/genomic junctional sequence by arbitrary-primed PCR and determined its sequence as described (Merrell *et al.*, 2002).

The primer pairs used for the arbitrary-primed PCR reactions (ARB1/MAG2F3 and ARB2/MAGF4) are listed in Table 4. DNA sequencing of arbitrary primed PCR products was performed by the W.M. Keck Facility at Yale University. Obtained nucleotide sequence was used to identify the precise site of the *magellan2* insertion in the TIGR4 genome sequence. The predicted protein sequence of each disrupted ORF was used to search the non-redundant NCBI protein database by BLASTP.

[00170] The site of the *magellan5* transposon insertions in the *rlrA* locus were first determined by PCR using either TNPAB-F or PFLA-R with the primer MarOUT, which anneals to either end of the mini-*mariner* transposons, followed by gel electrophoresis. Select PCR products were then purified with the Qiagen PCR purification and the DNA sequence of the transposon junction was determined using MarOUT by the Tufts University Core Sequencing Facility.

[00171] The protein sequences of sortase homologues were aligned using a ClustalW alignment in MacVector 7.0. Neighbor joining analysis based on the mean character distance was performed using PAUP* 4.0b10, and bootstrap values were calculated from 100 replicates.

Construction of mutant pools

[00172] To generate a large number of *S. pneumoniae* transposon insertion strains, chromosomal DNA was prepared from strain AC353, a streptomycin-resistant derivative of TIGR4 (Tettelin *et al.*, 2001), and mutagenized by *in vitro* transposition with *magellan2*. *Magellan2*, a mini-transposon derivative of *mariner*, inserts into the pneumococcal chromosome in a highly random manner (data not shown), requiring only a TA dinucleotide at the insertion site (Lampe *et al.*, 1996). Transposon mutagenesis was performed as described (Akerley *et al.*, 1998), except that 63 *magellan2* derivatives, each containing a unique 40 basepair (bp) signature tag were used. Following transposition, mutagenized DNA was transformed into naturally competent AC353 as described (Bricker and Camilli, 1999). Approximately 100 insertion strains were sequentially collected from each of the 63 *magellan2* derivatives into the wells of microtiter plates, resulting in 100 pools of 63 signature tagged insertion strains for STM screening.

Determination of colonization bottlenecks

[00173] In an animal, not all bacteria in an inoculum are able to overcome barriers that limit or restrict the number of bacteria that survive initially and begin to multiply. This phenomenon is commonly referred to as a 'colonization bottleneck' (although 'colonization' is an inaccurate term in cases like pneumococcal pneumonia that are acute infections and of limited duration. Since STM depends on all strains in the starting inoculum having an equal opportunity to infect a particular tissue, the population dynamics of AC353 in the murine lung were analyzed to determine whether a bottleneck existed. To address this, a group of female Swiss Webster adult mice were infected with a single STM pool of 63 unique strains at a dose of 10^5 CFU administered intranasally. At various times following inoculation, pairs of mice were euthanized and the number of CFU in the lungs from each animal was enumerated. After 12 h, the mice appeared healthy and no bacteria could be cultured from the lungs, suggesting that a severe bottleneck exists. In an attempt to circumvent this bottleneck, the inoculum was increased to 2×10^7 CFU and the number of CFU per mouse lung was determined as above. The larger inoculum resulted in the successful infection of all mice, as between $10^4 - 10^7$ CFU were recovered from all animals at all time points until the mice became moribund after approximately 48 h. Accordingly, all subsequent lung infection experiments were performed with an inoculum of 2×10^7 CFU.

[00174] A second variable assessed, was the potential for a limited number of strains to out-grow all others after initial adherence, thus preventing all 63 strains from being equally represented at a late stage of infection. To test this possibility, four mice were infected with a single STM pool, and the complexity of the bacterial populations remaining in the lungs of each mouse at a late stage of infection was determined and compared. The presence or absence of each strain in the lungs was assessed by recovery of the signature tags and hybridization to a master signature tag dot blot as described in the Materials and Methods. The full input pool strain complexity was maintained in all four mice after 44 h of infection, with the exception of a few strains absent from all mice, which represent *bona fide* attenuated strains (data not shown). Therefore, a pool complexity of 63 strains administered at 2×10^7

CFU/mouse results in all 63 strains having an equal opportunity to adhere and multiply in the mouse lung, and strains that fail to be recovered after 44 h are attenuated. Of note, the pool complexity used here is intermediate to that chosen for two prior STM screens in *S. pneumoniae*. Polissi *et al.* (Polissi *et al.*, 1998) used a pool complexity of 50 strains, mutagenized by plasmid insertion-duplication, to infect BALB/c mice in a murine model of pneumonia. In the other study, Lau *et al.* (Lau *et al.*, 2001) used a pool complexity of 96 strains, also mutagenized by plasmid insertion-duplication, to infect CD-1 mice in murine models of pneumonia and bacteremia.

[00175] Another potentially powerful application of STM is to track a population of bacteria that initially infect a single site, and subsequently spread to other sites in the animal. From such an analysis, it is possible to determine if a systemic infection is clonal or due to a larger founder population. If the latter was true for the case of *S. pneumoniae* spreading from the lung to the bloodstream, then two simultaneous STM screens could be conducted after intranasal inoculation; one in the lung and one in the blood. Instead, in mice infected with the same STM pool, the population of bacteria recovered from the bloodstream is randomly composed of only a few strains from the input inoculum (data not shown).

Selection of avirulent strains

[00176] To identify pneumococcal genes essential for lung infection in mice, an STM screened was done. In total, 100 pools comprising 6149 strains were screened. Each pool was inoculated into two mice and the bacteria were recovered after 44 h by plating homogenized lung tissue from each animal on Tryptic-Soy Agar blood plates. Chromosomal DNA was purified from the combined outputs from each animal and used as template DNA for the PCR amplification of the signature tags as described in the Materials and Methods. A similar procedure was followed to obtain signature tags from the input population of bacteria. The amplified signature tags were used to probe nitrocellulose dot blots containing all 63 tags and attenuated strains were identified by visually examining output blots for spots exhibiting a decreased hybridization signal compared to the input blot.

[00177] In the primary (1°) round of screening, 2101 candidate attenuated strains were identified in a non-stringent manner, i.e., all strains that gave a noticeably

reduced signal on the output blot compared to the input blot were selected. A more stringent secondary (2°) screen was then done on 2080 of these candidate attenuated strains. For this, the 2080 strains were assembled into smaller 2° pools of 40 strains and each pool was inoculated intranasally into two mice at 2 x 10^7 CFU. After amplifying the signature tags from the input and output bacteria, and hybridization to the master dot blot membranes, 1265 strains were selected that had a highly reduced output signal relative to the input signal. These virulence attenuated strains represent 20% of the total number of strains initially screened.

[00178] In order to narrow the focus of study, a re-examination of the 2° screen dot blot films was done to identify the subset of strains that were highly attenuated as determined by lack of any hybridization signal on the output blots. Through this analysis, 387 strains were identified, representing 6.3% of the total strains screened. The sites of transposon insertion in 337 of the 387 strains were determined by arbitrary-primed PCR and DNA sequencing of the *magellan2*/genome junctions essentially as described (Merrell *et al.*, 2002). Table 1 lists these 337 strains, along with information on the gene disrupted in each and a functional classification based on the TIGR4 genome sequence release (Tettelin *et al.*, 2001).

Quantification of virulence defects of selected mutants

[00179] To validate the results of the STM screen and to quantify the degree of virulence attenuation of individual strains, competition assays were done. The transposon insertion mutations from 17 of the 337 highly attenuated strains, including 12 strains with disruptions in putative transcriptional regulators, were backcrossed into the wild-type strain and tested by competition assay as follows. Each of the mutant strains was mixed with the wild-type strain at a 1:1 ratio, and inoculated intranasally into four or more mice and simultaneously into Todd Hewitt-Yeast extract (THY) broth. Bacteria were enumerated from the lungs at 44 h and after 5 h from THY broth by plating serial dilutions on media selective for both wild-type and test strains, and then replica plating the colonies to media selective for only the test strain. The *in vivo* competitive index (CI) was calculated by dividing the ratio of mutant to wild-type bacteria recovered from the lungs by the ratio of mutant to wild-type bacteria that were inoculated into each animal. Similarly, the *in vitro* CI was

calculated using THY broth cultures in order to assess general growth defects. The geometric means of the CIs for each strain are listed in Table 2; a mean CI of less than 1 indicates a defect in virulence (or growth *in vitro*) of the test strain. Of the 17 strains examined, 16 were attenuated for lung infection. Furthermore, 13 were outcompeted by greater than 10-fold, confirming that the selected strains are highly attenuated when tested against the wild-type parental strain. None of the 17 strains tested suffered gross defects in multiplication in broth *in vitro*. This analysis indicates that the majority of the 387 mutant strains identified by two successive rounds of STM screening are reproducibly attenuated in competition assays. Hence, the genes disrupted or whose expression is affected by the transposon insertion in these strains should be considered *bona fide* virulence factors.

Determination of virulence phenotypes in other infection models

[00180] After confirming the attenuation of virulence of the selected strains in the murine lung, we sought to determine if these strains have global virulence defects, or if they could be categorized into classes based on *in vivo* phenotypes in other animal assays. To this end, most of the set of confirmed lung attenuated strains, and many additional strains, were tested in competition assays in murine models of bacteremia and nasopharyngeal carriage. For each animal model, mutant and wild-type bacteria were prepared exactly as described for the lung infections, however, different inoculum sizes were utilized to assure proper representation of each strain. For the bacteremia model of infection, mice were inoculated with 10^6 CFU by intraperotineal injection (*i.p.*). Alternatively, 10^8 CFU were inoculated intranasally into mice using a small inoculum volume for the nasopharyngeal carriage model. After 20 h for bacteremia and 7 days for nasopharyngeal carriage, the bacteria were recovered from blood or nasopharyngeal washes respectively, and CIs were determined as described above.

[00181] Of the 24 strains that were tested in the bacteremia model, half were attenuated, albeit to varying degrees (Table 2). Four strains, two with insertions in transcriptional regulators (STM119 and STM210) and two with insertions in

biosynthetic genes (STM4 and STM208), had severe virulence defects of greater than 40-fold. Of the remaining eight attenuated strains, three had intermediate defects (14 to 19-fold) and five were only slightly attenuated. The remaining 12 strains that were tested against the parental strain were not attenuated. Together, these data show that a set of tissue specific virulence factors have been identified, including several putative transcriptional regulators.

[00182] Thirteen strains that were attenuated for lung infection were tested for their ability to colonize the nasopharynx in competition with the wild-type strain. As in the bacteremia model, not all of the strains that were attenuated for lung infection were attenuated for colonization of the nasopharynx. Eight of the tested strains were deficient at colonizing the nasopharynx, and all eight exhibited greater than a 10-fold colonization defect. Interestingly, these 8 strains had differing phenotypes in lung infection and bacteremia. Five were attenuated in all three of the animal models (Class IV), including one mutant in a putative transcriptional regulator (STM38) and a second in the response regulator of a two-component signal transduction system (STM185). The remaining 3 strains were not attenuated when tested in the bacteremia model, but were each severely outcompeted by the wild-type strain in the two animal models of infection that involve interactions with mucosal surfaces.

Identification of rlrA *and a sortase homologue required for infection*

[00183] Of the transcriptional regulators identified by STM and tested in additional animal models, one (STM64) putatively codes for a protein with 49% similarity to RofA and Nra from *S. pyogenes* (Fogg *et al.*, 1994; Podbielski *et al.*, 1999). This mutant strain was outcompeted by the parental strain in both the pneumonia and nasopharyngeal carriage models, but not the bacteremia model. A greater virulence defect was observed in the nasopharynx, where the *rlrA* strain was outcompeted 14-fold (Table 2). These findings suggest that RlrA regulates one or more genes that are important for the interaction of *S. pneumoniae* with mucosal surfaces in the respiratory tract.

[00184] In some strains of *S. pyogenes*, *rofA* regulates the expression of a divergently transcribed gene coding for Protein F, a factor that mediates attachment to fibronectin (Fogg *et al.*, 1994). The pneumococcal *rofA* homologue, herein named

rlrA, for *rofA*-like regulator, is divergently transcribed from six genes (SP0462 to SP0468), three of which have very weak homology to microbial surface components recognizing adhesive matrix molecules (MSCRAMMs), and three that have homology to sortases (Figure 1). Sortases are enzymes that catalyze the covalent linkage of a family of secreted proteins that contain an LPXTG (SEQ ID NO: 530) motif to the bacterial cell wall (Mazmanian *et al.*, 2001). Although these three *S. pneumoniae* sortases were apparent from the genome sequence (Pallen *et al.*, 2001; Tettelin *et al.*, 2001), no experimental data to characterize these has been reported. In work to be reported elsewhere, RlrA was found to regulate the transcription of these six genes, and thus the three putative MSCRAMMs have been named *rrgA, rrgB,* and *rrgC* for RlrA-regulated gene. In addition, based on strong homology to sortases, the existence of a fourth sortase elsewhere in the chromosome, and the results below, the three flanking genes have been named *srtB*, *srtC*, *srtD*. Interestingly, one of the mutants identified in the STM screen mapped to *srtD* (STM65 in Table 1), the terminal gene in the locus.

[00185] Given that the *rlrA* and *srtD* strains were attenuated for infection and colonization of mucosal surfaces, it was determined whether any of the other genes in the locus were also attenuated in the same mouse tissues. *Magellan5*, a mini-*mariner* transposon conferring spectinomycin resistance, was used to create insertions by *in vitro* transposition into two PCR products spanning the entire *rlrA* locus. The transposition products were transformed into naturally competent AC353, and 55 transposon insertion strains were selected. Each transposon insertion was coarsely mapped by PCR using a template specific primer and a transposon specific primer, and the junction sequence of the transposon/*magellan5* junctions in selected strains was determined by DNA sequencing. Transposon insertions were obtained throughout the locus, including numerous insertions in each gene in the locus, thus demonstrating that neither the sortase genes, nor the *rrg* genes are essential for growth of *S. pneumoniae in vitro*.

[00186] Each of the transposon insertion strains shown in figure 1, except for *rlrA*, was tested by competition assay in the murine model of pneumonia, as described above. The results from these experiments are shown in Figure 2A. Of the six strains tested, only the *rrgA* and the *srtD* strains had a virulence defect in the murine lung,

exhibiting a 18-fold and a 24-fold attenuation respectively (Figure 2A). The same six strains were also tested for colonization defects in the nasopharynx. In these assays, the *rrgA* strain was again attenuated, exhibiting a 5-fold defect, and the *srtB* strain had a modest 2-fold defect (Figure 2B). The other four genes were not required in either animal model. On the contrary, transposon insertions in *srtC* and *rrgC* resulted in a phenotype of hypercolonization in the nasopharynx. The basis by which these strains outcompete the wild-type strain was not investigated. In a further set of competition experiments, we tested the *rrgA*, *srtB*, and *srtD* strains for defects in survival during bacteremia. We found that none of these were attenuated in this model (Figure 2B). These data support the model whereby these factors are specific to the interaction of *S. pneumoniae* and mucosal surfaces.

Signature-tagged mutagenesis fails to isolate virulence attenuated acapsular strains

[00187] The extracellular polysaccharide capsule plays an absolute role in the pathogenesis of *S. pneumoniae*. The majority of the biosynthetic genes coding for the serotype 4 capsule (SP0337 to SP0353) appear to be organized into a single operon of approximately 15 kilobases (kb), representing about 0.7 % of the TIGR4 genome. In our STM screen and in two smaller scale STM screens in *S. pneumoniae* (Lau *et al.*, 2001; Polissi *et al.*, 1998) virulence attenuated acapsular mutants were not found. Negative results led us to the hypothesis that, either the capsule was not required by TIGR4 for lung infection of Swiss Webster mice, or that *magellan2* insertions into the capsular operon were deleterious to growth *in vitro*, and therefore could not be isolated.

[00188] To discern between these two possibilities, a 9.9 kb fragment of the capsule operon was amplified with primers cpsF1 and cpsR1 by PCR, and used as template DNA for *in vitro magellan2* transposon mutagenesis. The transposition products were transformed into wild-type bacteria and transposon insertion strains were selected on media containing Cm. No transformants were recovered after the standard 24 h of growth, however, after an additional 24 h a small number of transformants appeared. In contrast, in a parallel experiment, transposition into a non-essential 10 kb segment of the genome yielded a large number of transformants after 24 h of growth (data not shown). Thus the conclusion is that disruption of the TIGR4 capsular operon by *magellan2* is inhibitory to colony formation in the experimental

conditions used, and that the failure to recover attenuated acapsular mutants is most likely due to the low plating efficiency of these strains following transformation. The growth rate of an acapsular mutant, AC846 (see below), was found to be equivalent to the wild-type strain when grown individually or in co-culture experiments in THY broth showing that the low frequency selection of these strains is not simply due a general growth defect. Whether a similar phenomenon occurred in the other two STM screens remains unknown.

[00189] Several of the acapsular mutants isolated after 48 h of growth in the above experiment were confirmed by mapping of the *magellan2* insertion and by negative Quellung reactions (data not shown). An acapsular strain containing a disruption of *cps4E* (AC846, Table 3) was tested in competition assays in the murine lung and bacteremia models. In both instances the acapsular mutant was severely attenuated ($CI \leq 0.04$ and ≤ 0.001 for pneumonia and bacteremia models, respectively), confirming the importance of serotype 4 capsule in these animal models.

[00190] Novel insights into the pathogenesis of *S. pneumoniae* are likely to aid in the development of new antibiotic treatments and vaccines. Current knowledge of factors implicated in virulence have led to promising developments towards new protein based vaccines (Briles *et al.*, 2000a; Briles *et al.*, 2000b), however, an understanding of how most of these factors contribute to and function during infection is still lacking. In this study, the knowledge base of genes that are essential for virulence in a murine model of pneumonia has been greatly expanded by completing an STM screen; the third of its kind in this organism, but by far the most extensive.

[00191] Surprisingly, 20% of the 6147 strains screened by STM had a noticeable virulence defect. The large percentage of attenuated strains isolated in this screen is much higher than the 1 to 7 % observed in similar screens in other Gram-positive pathogens (Autret *et al.*, 2001; Jones *et al.*, 2000; Mei *et al.*, 1997). Of note, however, each of the previous pneumococcal STM screens identified approximately 10% of strains as attenuated (Lau *et al.*, 2001; Polissi *et al.*, 1998). The difference between prior pneumococcal STM screens and ours may result from more stringent cut-offs for selecting attenuated strains in the latter studies. Additionally, as has been suggested by others, it is conceivable that the use of a polar transposon mutagen may contribute to a higher percentage of attenuated strains within a library compared to

mutants isolated by plasmid insertion-duplication (Paton and Giammarinaro, 2001). Insertion of a polar transposon into a genetic locus not only disrupts the gene harboring the insertion, but also downstream genes that are cotranscribed with that gene. With plasmid insertion-duplication, not all insertions will result in a gene or operon null mutation, as for example plasmids containing either the 5' end of a gene or containing a promoter region will likely regenerate a wild-type copy of the same gene or promoter following recombination.

[00192] The three independent STM screens in *S. pneumoniae* have resulted in the combined screening of over 8500 strains for virulence defects in a number of different assays. Remarkably, there is very little overlap in the sets of genes that have been identified as essential virulence factors in each of these screens. Only 10 of the 231 unique genes identified here were also reported in previous *S. pneumoniae* STM screens (Lau *et al.*, 2001; Polissi *et al.*, 1998). The lack of significant overlap between the three screens is likely the result of two factors; 1) the number of *S. pneumoniae* genes that are crucial to survival *in vivo* is probably large, such that the combined STM screens have not approached saturation yet, and 2) the different mutagenesis strategies employed (transposon versus plasmid insertion-duplication) are responsible for mostly distinct sets of genes being disrupted. Regardless of the underlying causes, the lack of significant overlap between the three STM screens suggests that many additional factors linked to virulence remain to be identified.

[00193] To learn more about additional *in vivo* roles for some of the virulence genes identified herein, mutants were grouped based on their phenotypes in murine models of nasopharyngeal carriage and bacteremia. In total, 25 different strains were tested in multiple animal models using competition assays, and these are grouped by class in Table 2. From this a picture emerges of the tissue specificity that many *S. pneumonia* virulence factors play.

[00194] One striking feature of these classes is that many transcriptional regulators are found in each of the four classes, reinforcing the idea that tissue specific regulation of virulence factors is important for pneumococcal pathogenesis. Of the 16 putative transcriptional regulators identified in our screen, only two, *smrC* and SP2142 (STM119 and STM256) have been previously identified in *S. pneumoniae*, and thus most have unknown targets of regulation. Similarly, most of the two component signal transduction systems (TCSTS) in *S. pneumoniae* also have

unknown targets of regulation. In this screen, roles for five of the 13 *S. pneumoniae* TCSTSs are implicated in lung infection (Lange *et al.*, 1999; Throup *et al.*, 2000). In addition to the insertions isolated in *rr01*, *rr07*, *zmpR*, and *comD* (STM185, STM29, STM90, and STM281), the insertion in strain STM237 could potentially have polar effects on *hk11/rr11*. Four of these five strains were tested in competition assays. Three strains were attenuated in competition assays in the lung, but they each had different phenotypes in the bacteremia model; *zmpR* was not attenuated, *rr01* was attenuated 4-fold, and STM237 was attenuated 14-fold. Additionally, *rr01* was severely outcompeted by the wild-type strain during nasopharyngeal colonization, making it the only one of the three TCSTSs tested that was required in all three models.

[00195] Mutants in most of the TCSTSs have been tested previously for avirulent phenotypes. *comDE*, which is involved in the induction of natural competence (Pestova *et al.*, 1996), was previously shown to attenuate virulence in both lung infection and bacteremia (Bartilson *et al.*, 2001; Lau *et al.*, 2001). Two groups identified each of the *S. pneumoniae* TCSTS by sequence homology and tested mutant TCSTS strains in different animal models (Lange *et al.*, 1999; Throup *et al.*, 2000). Throup *et al.* (Throup *et al.*, 2000) used the respiratory tract infection (RTI) model to test TCSTS mutants for virulence defects, which employs single strain infections and relies on the titer of the mutant strain compared to the wild-type strain following 48 h of infection to determine the virulence phenotype. By this assay, mutations in *rr01* and *zmpR* each attenuated virulence, which is consistent with our findings, however, a mutation in *hk11/rr11* did not. Lange *et al.* (Lange *et al.*, 1999) examined mutant TCSTS strains for defects during systemic infection by determining the mean survival time of mice infected with each strain. None of the TCSTS mutant strains tested by this assay were attenuated, which conflicts with the observed phenotype of an *rr01* strain and STM237. These differences are best explained by the different animal assays used in each study. The previous experiments were done as single strain infections and used the survival of the animal to measure virulence defects, whereas our competition assays measure the ratio of the mutant and wild-type strain following coinfection to assess attenuation. Additionally, since our insertion in STM237 is upstream of the coding sequence for *hk11/rr11* in a putative ABC transporter, it is possible that this strain has a more severe phenotype than an insertion

in either *hk11* or *rr11* alone. In light of these findings and those of others (Bartilson *et al.*, 2001; Lau *et al.*, 2001; Throup *et al.*, 2000), it is interesting to speculate that these three two-component systems play important roles in the adaptation of *S. pneumoniae* to different host environments by sensing different extracellular signals that in turn result in differential virulence gene regulation.

[00196] The sequencing project of the TIGR4 strain identified a small number of loci that are not conserved in two other pneumococcal strains (Tettelin *et al.*, 2001). One such locus encodes *rlrA* (SP0461), a *rofA*-like transcriptional regulator, and six divergently transcribed genes including three putative MSCRAMM surface proteins. The present screen identified two genes in this locus, *rlrA* and *srtD*, and subsequently tested the other five genes for roles during infection. Three of the flanking genes, *rrgA*, *rrgB*, and *rrgC*, code for putative surface proteins that are homologous to MSCRAMM family members, and thus it is predicted that they may be involved in the attachment of *S. pneumoniae* to mucosal surfaces. Consistent with this hypothesis, the *rrgA* strain was attenuated in both the pneumonia and the nasopharynx carriage model, but not the bacteremia model.

[00197] In addition to having homology to MSCRAMMs, RrgA, RrgB, and RrgC have sorting signals that are characteristic of proteins that are anchored to the gram-positive cell wall by sortases (Mazmanian *et al.*, 2001). The sorting signal is composed of a C-terminal sequence consisting of an LPXTG motif (SEQ ID NO: 530), followed by a stretch of hydrophobic residues, and a series of charged residues (Schneewind *et al.*, 1993). RrgA, RrgB, and RrgC each have these characteristics, except that the leucine is replaced by a tyrosine, isoleucine and valine, respectively (Figure 1B). Since at least one cell-wall anchored protein (RrgA) is required for infection and colonization, one would predict that one or more sortases should also be required. Consistent with this hypothesis, we found that a mutation in *srtD* resulted in a severe defect in the ability of *S. pneumoniae* to infect the lung. Together with the observed phenotypes of other strains with mutations in the *rlrA* locus, these data suggest a specific role this locus in the interaction of *S. pneumoniae* with mucosal surfaces.

[00198] In *S. aureus*, it has been elegantly shown that sortases are transpeptidases that anchor target proteins by cleaving the peptide bond between the threonine and

glycine of the LPXTG (SEQ ID NO: 530) and covalently anchoring the threonine to the cell wall. Through the genomic sequence analysis of numerous Gram-positive organisms it is evident that multiple sortase paralogues are common within single strains, including TIGR4. In addition to *srtBCD*, the TIGR4 genomic sequence also contains a fourth sortase, *srtA*. SrtA is found in at least two other *S. pneumoniae* strains, R6 and D39, which do not contain *srtBCD* (Hoskins *et al.*, 2001). Given this, we hypothesize that *srtA* is the sortase orthologue common to all *S. pneumoniae* and other streptococci, whereas *srtBCD* may be 'specialized' sortases that have been acquired by only select strains to anchor specific proteins. To investigate this further, a phylogenetic tree was constructed using the four *S. pneumoniae* sortases and other sortase homologues from a number of Gram-positive bacteria (Figure 3). SrtA from both TIGR4 and R6 are found grouped in a clade with other SrtA orthologues including those from *S. gordonii*, *S. pyogenes*, and *S. suis*. SrtBCD, however, are rooted in two separate clades with non-SrtA sortase orthologues from *S. suis* and *S. pyogenes*. Together, these data indicate that sortases fall into at least two different groups, one group that contains the common sortases to many Gram-positive bacteria and a second group containing specialized sortases.

[00199] The role that multiple sortase paralogues play in protein anchoring has only been studied in two different species thus far. In *S. aureus* there are two known sortases, SrtA anchors the majority of the LPXTG (SEQ ID NO: 530) containing proteins, while SrtB has only been shown to anchor a single protein that contains an asparagine substituted for the leucine in the LPXTG motif (SEQ ID NO: 530). Furthermore, *srtB* is transcriptionally regulated in response to changing iron-conditions, rather than being expressed constitutively (Mazmanian *et al.*, 2002). In *S. suis*, five sortase homologues have been identified, and as in *S. aureus*, the majority of the anchored surface proteins are dependent upon a single sortase, SrtA. Given these findings, SrtA is susptected to anchor most LPXTG (SEQ ID NO: 530) containing proteins in *S. pneumoniae*, and the remaining sortases may then anchor a specific set of surface proteins in different environmental conditions in response to different environmental cues. It is tempting to speculate that the role for SrtB, SrtC, and SrtD proteins in TIGR4 is to anchor the (L)PXTG-motif (SEQ ID NO: 530) proteins RrgA, RrgB, and RrgC, which are coded by the genes flanking *srtBCD*.

EXAMPLE 2: Transcriptional Regulation in the *Streptococcus pneumoniae rlrA* Pathogenicity Islet by RlrA

MATERIALS AND METHODS

Bacterial strains, plasmids, and primers

[00200] The bacterial strains and plasmids used in this study are listed in Table 1. The parental strain for all *S. pneumoniae* genetic manipulations was AC353, a streptomycin resistant (Sm^R) derivative of TIGR4 (10). *S. pneumoniae* strains were grown in Todd-Hewitt broth plus 5% yeast extract (THY), and supplemented with 0.8% maltose when indicated. Unless otherwise stated, the antibiotic concentrations used in this study were as follows: Sm 100μg/ml, chloramphenicol (Cm) 4μg/ml, and spectinomycin (Spc) 200μg/ml for *S. pneumoniae*, and ampicillin (Ap) 100μg/ml, Cm 10μg/ml, Spc 100μg/ml for *E. coli*. Primers used in this study are listed in Table 2. Unless otherwise noted, all PCR reactions were performed in reaction buffer containing 1x *Taq* reaction buffer (Promega), 250μM dNTPs, 1μM of each primer, and a 10:1 mix of *Taq* and *Pfu* DNA polymerases. Reaction conditions consisted of 25 cycles of 95°C – 30s, 50 to 52°C – 30s, and 72°C – 30s/kb of DNA, followed by a 5 min post-dwell at 72°C.

Construction of an rlrA overexpressing strain

[00201] To construct a strain that expressed *rlrA* from an inducible promoter, the coding sequence of *rlrA* was introduced into the *S. pneumoniae* maltose locus downstream of *malM* (24). To this end, DNA fragments containing the 3′ end of the *malM* gene and the 5′ end of *malP* were PCR amplified from AC353 using the primer pairs MALFX/MALRP and MALPF2/MALPRP, respectively. Similarly, the *cat* gene, conferring Cm-resistance (Cm^R) in both *E. coli* and *S. pneumoniae*, was PCR amplified from pAC1000 with the primer set PCATF1/PCATR1 and the coding sequence of *rlrA* was PCR amplified from AC353 with the primer set RLRAFR/RLRARX. In the latter case, the Shine-Dalgarno sequence of the *S. pneumoniae rpoB* was engineered into the RLRAFR sequence to allow optimal translation efficiency of *rlrA* at the maltose locus. Each of these fragments were

subcloned separately into pCR-Script Amp SK(+) (Stratagene) and subsequently inserted into pAC1000, to generate pCH84. pAC1000 is a derivative of pEVP3 (3) that was created by PCR amplifying the pEVP3 vector backbone using the primer set PEVPF1/PEVPR1 to delete the promoterless *lacZ* gene in pEVP3. The resulting product was digested with *BamHI*, gel purified, and ligated overnight at 4°C. The final construct contains the 3′-*malM* sequence and the 5′-*malP* sequence flanking the *rlrA* coding sequence and the *cat* gene. To generate AC1278, the *S. pneumoniae* strain overexpressing *rlrA*, pCH84 was linearized by digestion with *XhoI*, and the gel-purified fragment was transformed into naturally competent AC353 as described (10). The double recombination event was selected by plating on Cm and confirmed by PCR and DNA sequencing.

[00202] *Ribonuclease protection assays (RPAs)*

[00203] Total RNA was isolated from 10mL exponential phase *S. pneumoniae* using the Qiagen RNAeasy kit according to the manufacturers recommendations (Qiagen). Template DNA for the generation of riboprobes was PCR amplified using the following primer sets: RLRAF2/RLRAR7, RRGAF3/RRGAR3, RRGBF2/RRGBR1, RRGCF2/RRGCR2, SRTBF2/SRTBR1, SRTCF2/SRTCR2, SRTDF2/SRTDR2, SRTAF1/SRTAR1, and RPOBF3/RPOBR3. The resulting products were purified using the QIAquick PCR purification kit, subsequently cloned into pGEM-T (Promega), and confirmed by PCR using both an SP6 or T7 primer and a primer specific to the cloned insert. These plasmids (AC1279 – AC1286, AC1293; Table 1) were used as templates for the generation of riboprobes as described (19). Synthesized probes were gel purified on a 4% denaturing polyacrylamide gel containing 7M urea. Ribonuclease protection assays were carried out as described by the manufacturer using the RPAII kit (Ambion). The protected fragments were visualized by exposing each gel to a phosphor imaging screen (Kodak) and analyzed using a Storm 860 scanner and IQMac V1.2 imaging software. The relative amount of each protected fragment in each assay was normalized to the amount of *rpoB* protected RNA in each lane.

[00204] *Northern blotting*

[00205] Northern blots were performed using the NorthernMax analysis kit (Ambion) exactly as described by the manufacturer using 5μg of total RNA. Riboprobes were synthesized as described above. Total RNA was separated on a 1%

formaldehyde agarose gel by electrophoresis and subsequently transferred to Hybond-N+ nitrocellulose membranes. Membranes were then probed with 10^6 cpm of gel purified riboprobe per mL of hybridization buffer, and washed as described (Ambion). Processed blots were exposed to a phosphor imaging screen (Kodak) and analyzed as described above.

[00206] *Primer extension and DNA sequencing*

[00207] Primer extension reactions were carried out using the AMV primer extension reverse transcriptase system (Promega). RNA was isolated from AC1278 as described above. A primer corresponding to 5' end of each coding sequence was end labeled with [γ-P^{32}]-ATP using T4 polynucleotidekinase (PNK) for 10 min. at 37°C. The primers used were: RLRAPE2, RRGAP2, RRGBPE, RRGCPE, SRTBPB, SRTCPE, and SRTDPE (Table 1). End labeled primers were annealed to total RNA extracted from AC1278 by incubation at 58°C for 20 minutes followed by cooling to room temperature for 10 min. AMV extension mixture was added to each annealed primer, and cDNA synthesis was carried out at 42°C for 30 min.

[00208] DNA fragments predicted to contain promoter regions in the islet were PCR amplified from AC353 using the following primer sets: RLRA2/RRGA2, RRGB2/RRGBR1, RRGC2/RRGBF2, SRTBP1/SRTBP2, and SRTCD1/SRTCD2. PCR products were purified using the Stratagene PCR purification kit according to the provided protocol (Stratagene) and purified products were subsequently cloned into pGEM-T (Promega) to generate plasmids AC1287, AC1288, AC1289, AC1290, and AC1291, respectively. DNA sequencing of *rlrA* pathogenicity islet promoter regions was performed using the Sequenase 2.0 DNA sequencing kit according to the manufacturers specifications (USB). Briefly, strains AC1287, AC1288, AC1289, AC1290, and AC1291 were grown in 4mL of LB broth and plasmid DNA was purified using the Qiagen mini plasmid prep system (Qiagen). Plasmid DNA was resuspended in 100µL of TE [10 mM Tris pH 8.0, 1 mM EDTA] and subsequently denatured by the addition of 25µL of 1N NaOH, 10mM EDTA and incubation at 37°C for 30 min. Single stranded DNA was ethanol precipitated by the addition of 1/10 vol of 3M sodium acetate (pH 5.2) and 2 vol 100% ethanol. Precipitated DNA was resuspended in 1X Sequenase reaction buffer and 60pmol of the appropriate primer was annealed by incubation at 37°C for 30 min. Sequencing reactions were performed by the addition of Sequenase 2.0 reaction mix containing [αS^{35}]-dATP and

incubation at room temperature for 5 min. Next, 3.5μL of each reaction was added to 2.5μL of each dideoxynucleotide at 37°C and the termination reaction was incubated for 5 min, at which time the reaction was stopped by the addition of stop solution.

[00209] Primer extension products and sequencing reactions were denatured for 10 min at 80°C prior to electrophoresis on a 5% polyacrylamide/7M urea sequencing gel (National Diagnostics). Gels were run at 45W, dried using the Biorad model 853 gel drying apparatus, and analyzed as above.

RlrA-His$_6$ purification

[00210] The predicted coding sequence of RlrA was PCR amplified from AC353 using primers RLRAC1/RLRAC2, subcloned into pGEM-T, and liberated by digestion with *NcoI* and *BglII*. The liberated fragment was ligated into similarly digested pQE60 to create AC1292. The resulting strain containing the coding sequence for RlrA with a C-terminal His$_6$ tag (SEQ ID NO: 550) was grown in 2 mL of LB containing Ap to an OD_{600} of 0.5 and expression of RlrA was induced by the addition of IPTG to 1mM for 2 h. Proper expression of RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) was assessed by separation of induced and uninduced culture cell extracts by SDS-PAGE and subsequently by Western blotting using anti-His$_6$ (His tag shown in SEQ ID NO: 550) antibody (Roche) according to the ECL Western blotting protocol (Amersham Pharmacia Biotech).

[00211] For the purification of RlrA-His$_6$ (His tag shown in SEQ ID NO: 550), 2L of AC1292 was grown as above and induced with IPTG for 2h. RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) was subsequently purified on a Ni^{2+}-NTA agarose column according to the manufacturers protocols (Qiagen). RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) containing fractions were combined and concentrated using Centricon centrifugation filters (Amicon) to a final concentration of 800 nM.

Gel shift assays of the rrgA-rlrA *promoter region*

[00212] Overlapping DNA fragments of the *rrgA-rlrA* intergenic region were amplified by PCR using the primer sets REGF1-AP3, IIR1-AP5, AP4-AP6, or IIR1-AP4 (AP7) and used in gel shift assays with RlrA-His$_6$ (His tag shown in SEQ ID NO: 550). In each experiment, 60pmol of a selected primer was end-labeled using T4 PNK (New England Biolabs) and [γ-P^{32}]-ATP (6000 Ci/mmol, 150 mCi/mL) for 30 minutes at 37°C. Labeled primers were ethanol precipitated with ammonium acetate twice, resuspended in 10μL of dH_20, and used in PCR reactions using pAC1287 as

template. Amplified products were separated on a 4% polyacrylamide gel, gel purified, and eluted overnight in gel shift elution buffer [0.5mM NH_4Ac, 10mM MgAc, 1mM EDTA, 0.1% SDS] at 37°C. Gel shift binding reactions were carried out using 5000 cpm of each probe with increasing concentrations of RlrA-His_6 (His tag shown in SEQ ID NO: 550) at 30°C for 15 min in gel shift binding buffer [20mM Tris (pH 8.0), 50 mM KCl, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT, 0.05% Nonidet P-40, 5% glycerol] supplemented with 1μg of (poly-dI·poly-dC)-(poly-dI·poly-dC) and bovine serum albumin as non-specific inhibitors. For the supershift experiments, binding reactions were performed as above, chilled on ice, and incubated with 0.5μg of anti-His_6 (His tag shown in SEQ ID NO: 550) antibody (Roche) for 30 min on ice. Reactions were subsequently separated on 5% non-denaturing polyacrylamide gel (Protogel; National Diagnostics) and visualized as described above.

DNaseI footprinting

[00213] DNaseI footprinting experiments were carried out using the gel-shift protocol with 2 x 10^4 cpm of each probe. Following protein binding, the concentration of $MgCl_2$ and $CaCl_2$ was adjusted to 5mM and 10mM and each reaction was incubated of DNaseI (0.5 to 2U) for 1 minute at room temperature. Reactions were stopped by the addition of stop solution (200mM NaCl, 30mM EDTA, 1% SDS) and the digested products were extracted with an equal volume of phenol and chloroform and subsequently ethanol precipitated. Precipitated DNA was resuspended in loading buffer (98% formamide, 10mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol) and separated on a 5% polyacrylamide/7M urea sequencing gel (National Diagnostics). Sequencing reactions of the footprinted region were performed as described above using primers specific to the region.

Determination of RlrA consensus binding sites

[00214] The consensus RlrA binding site was determined by PRETTY (GCG Software package) using the four RlrA binding sites determined by DNaseI footprinting. The resulting consensus sequence was used to query the complete TIGR4 genomic sequence using FINDPATTERNS (GCG Software package). The resulting sequences were analyzed to determine if the sequences were present in regions likely to contain *S. pneumoniae* promoters.

RESULTS

[00215] One putative transcriptional regulator identified by STM is RlrA (10), a homologue of RofA and Nra from *Streptococcus pyogenes* (6, 22). Through sequence analysis, *rlrA* has been shown to be one of seven genes in a pathogenicity islet of approximately 12 kb (see Figure 4)(10), which is not highly conserved in other *S. pneumoniae* strains (26). Of the six genes that are divergently transcribed from *rlrA*, three have homology to the LPXTG (SEQ ID NO: 530) family of cell wall anchored surface proteins (*rrgA*, *rrgB*, *rrgC*). RrgA, RrgB, and RrgC have C-terminal sorting signals that are characteristic of LPXTG (SEQ ID NO: 530) containing proteins, except that the leucine of the LPXTG (SEQ ID NO: 530) is deviant in each protein. RrgB and RrgC have conservative changes to isoleucine and valine respectively, whereas RrgA has a change to tyrosine. The C-terminal sorting signals predicts that these proteins are covalently anchored to the cell wall by sortases, which are transpeptidases found in most Gram-positive bacteria (15, 20). Interestingly, three of the four sortase homologues (*srtB*, *srtC*, and *srtD*) encoded in the TIGR4 genome lie within the *rlrA* pathogenicity islet, however, no proteins are known to be sorted by these sortases (see Figure 4) (10, 20, 26).

[00216] In addition to *rlrA*, *srtD* was also identified as an essential virulence gene through STM, and each was confirmed to be essential to the survival of *S. pneumoniae* during lung infection by testing strains with transposon insertions in each gene in competition assays against the wild-type parental strain (10). The *rlrA* gene was also found to be essential for colonization of the nasopharynx, but not bacteremia, whereas *srtD* was dispensable in both of these models (10). The generation of transposon insertion mutations in each of the remaining genes in the locus and subsequent analysis of each mutant strain in murine models of infection demonstrated that *rrgA* was also essential for colonization of the nasopharynx and lung infection, whereas *srtB* was essential for only for colonization of the nasopharynx (10).

[00217] Given the homology of RlrA to other Gram-positive transcriptional regulators, the organization of the islet, and the phenotypes of certain mutant strains in animal assays, we previously proposed a model of regulation in the *rlrA* pathogenicity islet in which RlrA positively regulates the transcription of each *rlrA*

pathogenicity islet gene. In the present work, we confirm this model by demonstrating that transcription of each gene in the islet is dependent upon RlrA. Furthermore, RlrA is shown to act at four different promoters within the islet at a consensus sequence that is found elsewhere in the *S. pneumoniae* chromosome, suggesting that although the *rlrA* pathogenicity islet may function autonomously at the level of both transcription and the protein secretion levels. there may be additional targets of regulation in the TIGR4 chromosome.

RlrA is required for wild-type levels of expression of each gene in the islet

[00218] To assess the effect of a *rlrA* mutation on the steady-state levels of mRNA for each gene in the islet, RPAs were performed using RNA isolated from wild-type AC353 or AC1213, a strain that harbors a transposon insertion in *rlrA*. Riboprobes specific to each islet gene, as well as to *rpoB*, were synthesized. The *rpoB* gene, which codes for the β-subunit of RNA polymerase, was used to probe the same RNA preparations as the *rlrA* islet probes to serve as a loading control in each experiment. In each case, the steady-state level of mRNA of each gene was decreased in the *rlrA* strain compared to the wild-type strain, albeit to differing degrees (Figure 5). The greatest decrease in message was observed for *rrgB* and *rrgC*, which were reduced by 10- and 11-fold, respectively. The *rrgA* message was only decreased by 2.5-fold in AC1213, suggesting that *rrgA* is transcribed from a promoter distinct from *rrgB* or *rrgC*. Lastly, *srtB*, *srtC*, and *srtD* mRNA was also dependent upon RlrA, with the observed decreases in mRNA levels being 6-, 7-, and 8-fold, respectively. Of note, the *srtB* probe protected three differently sized messages, suggesting the possibility that there are multiple transcriptional start sites within the sequence of the riboprobe.

[00219] To test a possible role of RlrA in the regulation of *srtA*, the fourth sortase homologue in *S. pneumoniae* that is unlinked from the *rlrA* islet, a riboprobe specific to the *srtA* coding sequence was generated. As above, an RPA was performed using total RNA harvested from either AC353 or AC1213. As shown in Figure 5B, there was no difference in the amount of protected *srtA* message in either strain, indicating that *srtA* transcription occurs independently of RlrA.

RlrA is autoregulatory

[00220] In *S. pyogenes*, RofA positively regulates its own expression (5). To investigate the possibility that RlrA functions in a similar manner, a merodiploid strain that overexpressed *rlrA* from an inducible promoter was constructed (AC1278). This strain contained two copies of *rlrA*; one present in the *rlrA* pathogenicity islet and a second copy integrated into the maltose utilization operon downstream of *malM* (24). In the latter case, expression of *rlrA* was under the control of the *malM* promoter (P_{malM}), and thus its expression was inducible by the addition of maltose to the growth media (1). In addition, the Shine-Dalgarno site of *rpoB* was engineered into the *rlrA* construct upstream of the *rlrA* initiation codon to assure maximal translation efficiency of RlrA from the maltose utilization locus.

[00221] To determine if overexpression of *rlrA* (from P_{malM}) activated transcription from the native *rlrA* promoter (P_{rlrA}), RPAs were performed using a single riboprobe to *rlrA* that differentiated between the two transcripts. The riboprobe was completely complementary to the P_{rlrA} transcript, as it overlapped the coding sequence of *rlrA* and the 5' untranslated mRNA, resulting in a 409 bp protected band. Alternatively, the *rlrA* riboprobe was only partially complementary to the P_{malM} transcript, and resulted in a smaller protected fragment since the sequence upstream of the *rlrA* coding sequence in this locus is different from that in the *rlrA* pathogenicity islet. Due to these differences, the two differently sized protected messages detected with the same riboprobe were used to assess the quantity of steady-state mRNA from either of these promoters.

[00222] As shown in Figure 5C, an increase in the amount of *rlrA* mRNA initiated from P_{rlrA} was observed in strain AC1278 compared to AC353 when each strain was grown in the absence of maltose. The increase observed in the absence of inducer compared to AC353 was due to the fact that AC1278 contained two copies of *rlrA* and transcription from P_{malM} is not completely repressed during growth in THY. A 6-fold increase in expression from P_{rlrA} was observed in strain AC1278 compared to AC353 when each strain was grown in the presence of maltose. No increase in *rlrA* expression was observed when AC353 was grown in maltose compared to the same strain grown in THY, confirming that the increase in *rlrA* expression in AC1278 is not due to simply to growth in the presence of maltose. Together these data show that RlrA is autoregulatory in addition to activating the expression of the 6 other genes in the *rlrA* pathogenicity islet.

Transcription in the rlrA pathogenicity islet initiates at four different promoters

[00223] The finding that AC1213 (*rlrA::magellan2*) exhibited decreased levels of steady-state mRNA of different genes in the islet by differing levels led to the hypothesis that RlrA acts at numerous sites within the locus. To identify sites of transcription initiation, and thus sites of potential RlrA activity, a primer specific to each gene is the locus was synthesized and used in primer extension analysis. By this method, transcription initiation sites upstream of the *rlrA*, *rrgA*, *rrgB*, and *srtB* were identified (Figure 6). By analyzing the sequences upstream of the predicted transcriptional start sites, σ^{70} consensus –10 and –35 sequences were identified for *rlrA*, and an extended –10 sequence (25) was found for *rrgA* and *rrgB*, however, no such sequences were found for the *srtB* promoter (Figure 6A and 6B). These results support the model that there are multiple promoters within the islet, and thus multiple sites at which RlrA acts.

[00224] Efforts to identify transcriptional start sites upstream of the predicted open reading frames of *rrgC*, *srtC*, and *srtD* proved unsuccessful. This suggested that each of these genes was transcribed from a distal promoter and that each was cotranscribed with an upstream gene(s). To test this, Northern blots were carried out using total RNA extracted from either AC1278 or AC1213 grown in THY-maltose Using the same riboprobes that were used for RPAs, we found that the *rrgC* probe hybridized to an mRNA of approximately 3.8 kb, the predicted size of a mRNA that would include both *rrgC* and *rrgB*. In support of this, a Northern blot probed with *rrgB* indicated a message of the same size (Figure 7A, lanes 1). No message corresponding to *rrgC* or *rrgB* could be detected in the *rlrA* mutant strain consistent with the RPA data that transcription of both genes is dependent upon RlrA (Figure 7A, lanes 2).

[00225] When the same RNA preps were probed with riboprobes complementary to *srtB*, *srtC*, and *srtD*, an mRNA of approximately 2.7 kb was detected with all three probes in the AC1278 background (Figure 7B). A similar sized message was detected in AC1213 at sharply decreased levels, although the same amount of RNA was loaded in each lane as determined by the quantity of rRNA on ethidium bromide stained agarose gels (data not shown). An additional mRNA species of approximately 3.7 kb was also detected using the *srtB* probe that was not found with the *srtC* or *srtD* probe.

Given the size of the message, its dependence on RlrA, and the position of the *srtB* riboprobe (which is predicted to overlap the *rrgBC* message), this mRNA is predicted to be the *rrgBC* message that terminates immediately upstream of *srtB* coding sequence.

RlrA-His$_6$ acts at the rrgA and rlrA promoters

[00226] To determine if RlrA directly acts at one or more of the promoters in the *rlrA* pathogenicity islet, a C-terminally His$_6$-tagged (SEQ ID NO: 550) version of RlrA was purified from *E. coli*. To test if RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) was able to bind to *rlrA* pathogenicity islet promoter sequences, the noncoding sequence between *rrgA* and *rlrA* was amplified by PCR using the primer set REGF1/IIR1 using an end-labeled REGF1 primer (Figure 8A). The resulting fragment was incubated with purified RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) and separated on a nondenaturing polyacrylamide gel. In this gel-shift assay, RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) retarded the mobility of the probe evinced by the presence of multiple species that migrated more slowly on the gel than the probe alone (data not shown). This results show that RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) retains DNA binding activity, and indicates that it binds to multiple sequences between *rrgA* and *rlrA*.

[00227] To more finely map the regions that the purified protein bound to, smaller overlapping fragments of same region of DNA were generated by PCR and used in gel shift experiments (Figure 8B). When incubated with the AP4 fragment, RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) retarded the mobility of the probe, resulting in a single band that increased in intensity as the concentration of protein was increased (35% mobility shift at 4nM and 70% mobility shift at 16nM). A similar result was observed when RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) was incubated with the AP5 fragment, however, as the concentrations of protein were increased, two retarded species were observed (50% mobility shift at 4nM RlrA-His$_6$ (His tag shown in SEQ ID NO: 550)). With both AP4 and AP5 probes, as well as, with the AP3 probe that spans the intergenic region downstream of the *rrgA* transcriptional start site, a retarded species running at the top of the gel was observed at high protein concentrations (RlrA-His$_6$ (His tag shown in SEQ ID NO: 550)> 130nM). We believe that this band is the result of nonspecific binding of RlrA-His$_6$ (His tag shown in SEQ ID NO: 550) at high concentrations, an idea supported by the binding of RlrA-His$_6$

(His tag shown in SEQ ID NO: 550) to non-promoter regions of *rrgA* (AP3) and to two other *S. pneumoniae* promoters that are not regulated by RlrA (data not shown).

[00228] To confirm that the retarded mobility of the probe was due to the binding of RlrA-His_6 (His tag shown in SEQ ID NO: 550) and not a contaminating species in the purified protein prep, anti-His_6 (His tag shown in SEQ ID NO: 550) antibody was added at the conclusion of the binding reaction to super-shift RlrA-His_6 (His tag shown in SEQ ID NO: 550) specific species. Figure 8C shows that incubation of RlrA-His_6 (His tag shown in SEQ ID NO: 550) bound complexes with anti-His_6 (His tag shown in SEQ ID NO: 550) antibody results in the appearance of a third complex migrating higher on the gel. This tertiary complex demonstrates that it is indeed RlrA-His_6 (His tag shown in SEQ ID NO: 550) that is bound to the AP4 and AP5 probes. Together, these data suggest that RlrA-His_6 (His tag shown in SEQ ID NO: 550) specifically binds to three distinct sites between the *rrgA* and *rlrA* transcription initiation sites resulting the activation of transcription from both the *rrgA* and *rlrA* promoters.

Determination of RlrA-His_6 binding sites

[00229] DNaseI footprinting was used to precisely map the sites of RlrA binding in the *rrgA* and *rlrA* promoter regions. RlrA-His_6 (His tag shown in SEQ ID NO: 550) was incubated with the AP7 fragment as described above, and the resulting bound complexes were subjected to DNaseI digestion. Consistent with the findings of the gel-shift experiments, RlrA-His_6 (His tag shown in SEQ ID NO: 550) protected three discrete regions of DNA (Figure 9A). Two of these regions were present within 80 bp of the *rlrA* transcriptional start site (–34 to –49; –53 to –82) and a larger third region was present within 70 bp of the *rrgA* transcriptional start site [–36 to –76; (Figure 9A and B)], which we believe constitutes two binding sites that are similar in arrangement to the *rlrA* binding sites. When the complementary strand of DNA was end-labeled and used in the same assay, the same binding patterns were identified for both the *rlrA* and *rrgA* promoters (data not shown). Alignment of the protected regions revealed that RlrA binds to AT rich regions close to the RNA polymerase binding site and transcriptional start site of each gene (Figure 9B). The four identified binding sites were aligned and a 15 bp RlrA consensus binding site was determined as RY(T/G)TTTTTR(T/A)(C/A)RA (SEQ ID NO:536). The resulting AT rich sequence was used to query the TIGR4 genome sequence for additional RlrA binding sites.

This search resulted in the identification of 153 sequences, 27 of which were present in putative promoter regions, and 14 that are within 15 bp of the –35 sequence. These data suggest that RlrA may regulate additional genes outside of the *rlrA* pathogenicity islet.

[00230] The *rlrA* gene was initially identified as an essential gene for the colonization of *S. pneumoniae* in the murine nasopharynx and for its ability to infect the murine lung (10). In addition, several genes that are divergently transcribed from *rlrA* and lie within a 12 kb stretch of DNA that is flanked by two insertion elements have also been shown to be essential for either or both of these two models (Figure 4). The *rrgA* gene codes for a predicted cell wall anchored protein of the LPXTG (SEQ ID NO: 530) family of Gram-positive surface proteins (10, 20). The LPXTG motif (SEQ ID NO: 530) is part of a larger C-terminal sorting signal that targets the protein to a specific pathway that ultimately covalently anchors the protein to the cell wall (16). The enzymes that anchor proteins to the cell wall in this manner are called sortases. Sortases are transpeptidases that cleave between the threonine and glycine of the LPXTG motif (SEQ ID NO: 530) resulting in the anchoring of the N-terminal half of the protein by a peptide bond between the threonine and the cell wall. Interestingly, also divergently transcribed from *rlrA* are three sortase homologues, *srtBCD* (Figure 4). Two of these three genes have been shown to have a role during *in vivo* survival; *srtD* is essential for lung infection and *srtB* is essential for colonization of the nasopharynx (10).

[00231] RlrA exhibits amino acid sequence similarity to a number of *S. pyogenes* transcriptional regulators, including RofA and Nra, a positive and negative regulator, respectively. Both RofA and Nra regulate their own expression, as well as, a number of different surface proteins that interact with eukaryotic extracellular matrices, and thus are important to the pathogenesis of *S. pyogenes* (6, 8, 22). In each case, the gene divergently transcribed from the regulator is one target of regulation.

[00232] To determine if RlrA was a regulator of neighboring genes and of its own transcription, RPAs were used to measure the steady-state levels of transcription of each gene in the *rlrA* pathogenicity islet. We found that RlrA positively regulates the transcription of each gene (Figure 5). The fold decrease in each message was determined in the *rlrA* strain compared to the wild-type strain. From this analysis, *rlrA* dependent expression fell into three categories; expression of the *rrgA* gene was

only slightly affected, *srtBCD* expression was decreased to an intermediate level, and *rrgBC* expression was drastically reduced. The role of RlrA in its own expression was analyzed using a merodiploid strain that expressed *rlrA* from the *malM* promoter, allowing inducible expression in the presence of maltose. Analysis of this strain revealed that RlrA positively regulates its own transcription.

[00233] The different levels of expression of each gene in the islet suggest that RlrA regulates multiple promoters within the islet. Indeed, using primer extension analysis, we mapped transcripts initiating upstream of *rlrA*, *rrgA*, *rrgB*, and *srtB*. A consensus σ^{70} –35 and –10 binding site was found upstream of *rlrA*, indicating that *rlrA* is expressed constitutively, but may also be subject to positive autoregulation by increased RlrA levels under unknown conditions. In contrast, in three cases, *rrgA*, *rrgB*, and *srtB* σ^{70} consensus –35 boxes could not be identified upstream of the transcriptional start sites, however, extended –10 sequences were identified for *rrgA* and *rrgB*. Previous studies on other *S. pneumoniae* promoters have shown that consensus sequences cannot always be found within DNA fragments with known promoter activity (25). It is conceivable that genes such as those in the *rlrA* pathogenicity islet are transcribed by alternative σ factors, such as ComX, which regulates a subset of competence induced genes (14). Comparison of the sequences upstream of the *srtB* promoter to the consensus *comX* box did not reveal an obvious binding site (21), indicating this promoter is ComX independent and may be transcribed using an unknown sigma factor that is aided by RlrA binding. Alternatively, RlrA may enhance transcription by stimulating binding of σ^{70}-RNAP holoenzyme to the poor –35 elements in the *rrgA*, *rrgB*, and *srtB* promoters.

[00234] The identification of multiple promoters that are regulated by RlrA indicated that RlrA must bind multiple sites within the islet to regulate gene expression. This was indeed shown to be the case by gel shift analyses and DNaseI footprinting. In these experiments, RlrA was demonstrated to directly bind to four sites within the *rlrA-rrgA* intergenic region; two sites upstream of *rlrA* and two sites upstream of *rrgA*.. In each case, there is a smaller RlrA binding site near the transcriptional start site and a larger binding site at a more distal location. A 15 bp consensus sequence in present in all four sites, and we propose that it is this sequence that is bound directly by RlrA. It is curious that the smaller site in each promoter overlaps the –35 sequence in each promoter, which is expected to be bound by RNA

polymerase. As mentioned above, a consensus σ^{70} –35 promoter sequence could be identified in the *rlrA* promoter, but not the *rrgA*, *rrgB*, or *srtB* promoter. Thus, these data suggest that RlrA may compete with σ^{70} for the smaller binding site in the *rlrA* promoter, possibly when RlrA is expressed at high levels, resulting in repression of RlrA expression.

[00235] An interesting aspect to the biology of the *rlrA* pathogenicity islet is that it is not conserved among all pneumococcal serotypes (11, 26). Therefore, this islet may require a means of autonomous regulation as we demonstrate here. On the other hand, it may seem unlikely that RlrA would regulate chromosomal genes outside the islet. However, we identified a number of putative targets of *rlrA* regulation, outside the islet and scattered throughout the genome. It will be interesting to analyze these loci to see if they are indeed regulated by RlrA.

[00236] The *srtBCD* genes represent three of the four sortase homologues in the TIGR4 *S. pneumoniae* genome. The presence of multiple sortase homologues is a common occurrence in Gram-positive bacteria genomes. The role of sortases in the anchoring of surface proteins important for the pathogenicity of various organisms is well documented. To our knowledge, however, prior to the finding that *srtBCD* are regulated by RlrA, only one other sortase has been shown to be regulated at the transcriptional level (17). The finding here that three of the four pneumococcal sortases are under the coordinate regulation of a single regulator suggests that RlrA may indirectly regulate the expression of numerous cell wall anchored proteins by controlling sortase expression from a single promoter. It remains formally possible that the multiple sortase homologues in the *rlrA* pathogenicity islet do not have substrates that lie outside of the islet. In this case, the role of SrtB, SrtC, and SrtD may be to specifically anchor one or more of the Rrg proteins to the cell wall. This would add a second, post-translational level of autogenous regulation to the *rlrA* pathogenicity islet.

Table 1. *Streptococcus pneumoniae* genes essential for lung infection

TIGR Designation	Strain Name	Gene Name[a]	Homologue[b]	Description	Reference[c]
Amino acid biosynthesis					
SP0445	STM60, STM61, STM62		*ilvB*	Isoleucine and valine biosynthesis	
SP0856	STM110		*ilvE*	Isoleucine and valine biosynthesis	
SP1377	STM165		*aroD*	Aromatic amino acid biosynthesis	
SP1544	STM182	*aspC*		Glutamate biosynthesis	
SP1815	STM207		*trpD*	Tryptophan biosynthesis	
SP1816	STM208		*trpG*	Tryptophan biosynthesis	
SP1817	STM209		*trpE*	Tryptophan biosynthesis	
SP1970	STM233		*asnA*	Asparigine biosynthesis	
SP2210	STM277, STM278	*cysM*		Cysteine biosynthesis	
Biosynthesis of cofactors, prosthetic groups, and carriers					
SP0177	STM32	*ribE*		Riboflavin biosynthesis	
SP0586	STM77			Folic acid biosynthesis	
SP0726	STM98	*thiD*		Thiamin biosynthesis	
SP2095	STM248			Folic acid biosynthesis	
Cell envelope					
SP0057	STM5	*strH*		β-N-acetylglucosaminidase	
SP0102	STM17		*wbgW*	Glycosyl transferase	
SP0136	STM22		BH3713	Glycosyl transferase, family 2	
SP1529	STM180			Putative polysaccharide biosynthesis protein	
SP1770	STM197			Glycosyl transferase	
SP1771	STM198			Glycosyl transferase	
SP1772	STM199, STM200, STM201		*hsa*	Cell wall anchored protein	
SP2017	STM238, STM239		SPY0196	Membrane protein	
SP2098	STM249			Membrane protein	
SP2136	STM329	*pcpA*		Choline binding protein	(Sanchez-Beato *et al*., 1998)
SP2145	STM260, STM261	*smuD*		Cell wall surface anchor family	(Lau *et al*., 2001)
SP2176	STM271, STM272, STM273		*dltA*	Lipidteichoic acid biosynthesis	
Cellular Processes					
SP0071	STM9, STM10, STM11			IgA1 protease	(Polissi *et al*., 1998)
SP0117	STM19	*pspA*		Choline binding protein	(Berry and Paton, 2000; Hammerschmidt *et al*., 1999; Hollingshead *et al*., 2000; Yother and White, 1994)
SP0268	STM44, STM45, STM46	*spuA*		Pullulanase	(Bongaerts *et al*., 2000; Zysk *et al*., 2000)
SP0314	STM52			Hyaluronate lyase	(Berry *et al*., 1994)
SP0377	STM57	*cbpC*		Choline binding protein	
SP0498	STM72			N-endo-b-N-acetylglucosamidase	(Zysk *et al*., 2000)
SP0641	STM86	*prtA*		Serine proteinase	(Wizemann *et al*., 2001; Zysk *et al*., 2000)
SP0648	STM88	*bgaA*		β-galactosidase	(Zysk *et al*., 2000)
SP0690	STM95	*divIB*		Cell division protein	
SP0766	STM101	*sodA*		Superoxide dismutase	(Yesilkaya *et al*., 2000)
SP0966	STM122	*pvaA*		Pneumococcal vaccine antigen A	(Wizemann *et al*., 2001)
SP0978	STM123, STM124	*coiA*		Competence induced gene	
SP1154	STM146	*iga*		IgA1 Protease	(Poulsen *et al*., 1998)
SP1645	STM187		*relA*	GTP pyrophosphokinase	

SP1889	STM220	*amiD*		Oligopeptide permease	(Cundell *et al*., 1995)
SP1890	STM221	*amiC*		Oligopeptide permease	(Cundell *et al*., 1995)
SP1891	STM222, STM223, STM224, STM225	*amiA*		Oligopeptide permease	(Cundell *et al*., 1995)
SP1923	STM226, STM227	*ply*		Pneumolysin O	(Berry and Paton, 2000; Walker *et al*., 1987)
SP1964	STM232	*endA*		DNA uptake nuclease	
SP1976	STM234	*pflA*		Pyruvate formate lyase activating enzyme	
SP2052	STM242	*cglB*		Competence induced gene	
SP2076	STM244	*hexA*		DNA mismatch repair protein	
SP2190	STM275	*cbpA*		Choline binding protein	(Rosenow *et al*., 1997)
SP2201	STM276	*cbpD*		Choline binding protein	(Gosink *et al*., 2000)
Central intemediary metabolism					
SP0253	STM40	*gldA*		Glycerol dehydrogenase	
DNA metabolism					
SP0023	STM1		*radA*	DNA repair	
SP0274	STM47, STM48		*polC*	DNA replication	
SP0510	STM73		*ecoAI*	Restriction modification	
SP0886	STM112, STM113		*hsdM*	Methyltransferase	
SP0887	STM114, STM115		MJ1218	Restriction modification	
SP0892	STM116, STM117		*ecoKI*	Restriction modification	
SP1040	STM134			Site-specific recombinase	
SP1202	STM153	*recN*		DNA repair	
SP1431	STM174		*M.XbaI*	Restriction modification	
Energy metabolism					
SP0240	STM35		*yjfF*	Phosphoglycerate mutase family protein	
SP0251	STM39	*smmF*		Fermentation	(Lau *et al*., 2001)
SP0265	STM42		*bglA.2*	Glycosyl hydrolase, Family 1	
SP0312	STM51		*xylS*	Glycosyl hydrolase, family 31	
SP0829	STM108		*deoB*	Purine salvage pathway	
SP0916	STM118	*cad*		Lysine decarboxylase	
SP1118	STM140			Pullulanase	
SP1121	STM141	*glgB*		Glycogen biosynthesis	
SP1193	STM152	*lacA*		Lactose catabolism	
SP1382	STM168	*amy*		α-amylase	
SP1855	STM213		*ypjA*	Fermentation	
SP1898	STM225	*aga*		a-galactosidase	
SP1998	STM235		*asnB*	L-asparingase	
SP2128	STM254		*tkt*	Transketolase, Pentose phosphate pathway	
SP2167	STM267		*fucK*	L-fuculose kinase	
Fatty acid and lipid metabolism					
SP0199	STM34	*cls*		Lipid biosynthesis	
SP0614	STM80	*estA*		Tributyrin esterase	
Hypothetical proetins					
SP0095	STM14		SPY0915	Conserved hypothetical	
SP0100	STM15, STM16		SPY2172	Conserved hypothetical	
SP0110	STM18			Hypothetical	
SP0145	STM24, STM25		*dsg*	Conserved hypothetical	
SP0146	STM26		*yqfD*	Conserved hypothetical	
SP0157	STM30			Hypothetical	
SP0160	STM31		*sdhB*	Conserved hypothetcial	
SP0198	STM33		SA1341	Conserved hypothetcial	
SP0298	STM49		Hi1038	Conserved hypothetical	
SP0332	STM54		*orfB*	Hypothetical	
SP0385	STM58		SPY1623	Conserved hypothetical	
SP0454	STM63		MJ1577	Conserved hypothetical	

SP0492	STM69			Hypothetical	
SP0595	STM78			Hypothetical	
SP0633	STM84, STM85			Hypothetical	
SP0663	STM91		spy1898	Conserved hypothetical	
SP0686	STM94		*orfA*	Conserved hypothetical	
SP0719	STM96		*ykoE*	Conserved hypothetical	
SP0728	STM99			Hypothetical	
SP0767	STM102		SPY1354	Conserved hypothetical	
SP0774	STM103			Hypothetical	
SP0785	STM104		SPY0836	Conserved hypothetical	
SP0789	STM105		*yveF*	Conserved hypothetical	
SP0939	STM120		PM0632	Conserved hypothetical	
SP0986	STM126, STM127		BH2069	Conserved hypothetical	
SP1003	STM129	*phtD*		Conserved hypothetical	(Adamou *et al*., 2001)
SP1045	STM135		SA1714	Conserved hypothetical	
SP1111	STM137		BH1678	Conserved hypothetical	
SP1127	STM142		SPY0729	Conserved hypothetical	
SP1143	STM143, STM144		HI0660	Conserved hypothetical	
SP1153	STM145			Hypothetical	
SP1174	STM148, STM149, STM150	*phtB*		Conserved hypothetical	(Adamou *et al*., 2001)
SP1175	STM151	*phtA*		Conserved hypothetical	(Adamou *et al*., 2001)
SP1281	STM155		*blpT*	Hypothetical	
SP1344	STM164		SC5A7.31	Conserved hypothetical	
SP1378	STM166, STM167		SPY0808	Conserved hypothetical	
SP1405	STM173		SPY1249	Conserved hypothetical	
SP1518	STM179		SPY0348	Conserved hypothetical	
SP1652	STM189		SPY1255	Conserved hypothetical	
SP1654	STM190	*smuB*		Conserved hypothetical	(Lau *et al*., 2001)
SP1706	STM191			Hypothetical	
SP1760	STM 194, STM195, STM196			Conserved hypothetical	
SP1779	STM202			Hypothetical	
SP1793	STM205			Hypothetical	
SP1879	STM219		SPY0369	Conserved hypothetical	
SP1952	STM230			Hypothetical	
SP1956	STM231		*orfA*	Hypothetical	
SP2002	STM236, STM237		SA1157	Conserved hypothetical	
SP2039	STM241		*sapR*	Conserved hypothetical	
SP2105	STM251			Hypothetical	
SP2143	STM257, STM258, STM259		*ypdB*	Conserved hypothetical	
SP2146	STM262		XF0106	Conserved hypothetical	
SP2159	STM263, STM264			Hypothetical	
SP2182	STM274			Hypothetical	
Other catergories (46 Strains)					
Protein fate					
SP0150	STM27, STM28		SSO2737	Peptidase M20/M25/M40 Family	
SP0338	STM55, STM56		*clpL*	ATP-dependent Clp proteinase	
SP0468	STM65	*srtD*		Sortase-like protein	
SP0664	STM92	*zmpB*		Metalloprotease	(Polissi *et al*., 1998)
SP0797	STM106	*pepN*		Aminopeptidase N	
SP0979	STM125	*pepF*		Oligopeptidase F	
SP1343	STM163			Prolyl family oligopeptidase	
SP1538	STM181		SPY1619	Protein folding and stabilization	
SP1591	STM184	*pepQ*		Proline dipeptidase	

SP1780	STM203, STM204		*pepF*	Oligoendopeptidase F	
SP2060	STM243	*pcp*		Pyrrolidone-carboxylate peptidase	
SP2239	STM282		*htrA*	Serine protease	
Protein synthesis					
SP0128	STM21		SPY1873	Alanine acetyltransferase	
SP0254	STM41	*leuS*		Leucyl-tRNA synthetase	
SP1029	STM132		SPY1346	RNA methyltransferase	
Purines, pyrimidines, nucleosides, and nucleotides					
SP0045	STM2	*purL*		Purine biosynthesis	(Polissi *et al*., 1998)
SP0050	STM4		*purH*	Purine biosynthesis	
SP0494	STM70, STM71	*pyrG*		CTP synthase	
SP0842	STM109	*pyn*		Pyrimidine-nucleoside phosphorylase	
SP1018	STM130	*tdk*		Thymidine biosynthesis	
SP1847	STM211	*xpt*		Purine salvage pathway	
Regulatory functions					
SP0141	STM23		*mutR*	Transcription factor	
SP0246	STM37		SPY2054	Transcription factor	
SP0247	STM38		SPY2053	Transcription factor	
SP0306	STM50		*yjdC*	BglG family antiterminator	
SP0461	STM64	*rlrA*		Transcription factor	
SP0807	STM107		SPY0728	FtsZ regulator	
SP0927	STM119	*smrC*		Transcription factor	(Lau *et al*., 2001)
SP1115	STM139		*mutR*	Transcription factor	
SP1278	STM154		*pyrR*	Transcription factor	
SP1433	STM175, STM330			Transcription factor	
SP1800	STM206		*dmgB*	Transcription factor	
SP1830	STM210		*phoU*	Transcription factor	
SP1854	STM212	*galR*		Transcription factor	
SP1856	STM214, STM215		*rmeA*	Transcription factor	
SP2131	STM255		SPY0952	Transcription factor	
SP2142	STM256		SPY1596	Transcription factor, ROK family	(Polissi *et al*., 1998)
Signal transduction					
SP0063	STM6, STM7, STM8		*ptnD*	Sorbose family PTS System	
SP0156	STM29	*rr07*		Response regulator	(Lange *et al*., 1999; Throup *et al*., 2000)
SP0396	STM59	*mtlF*		Mannitol PTS system	
SP0474	STM66			Cellobiose family PTS System	
SP0478	STM67	*lacE*		Lactose PTS System	
SP0645	STM87		SPY1711	Galacticol family PTS system	
SP0661	STM90	*cmpR*		Response regulator	
SP0877	STM111		SPY0855	Fructose PTS system	
SP1633	STM185	*rr01*		Response regulator	(Lange *et al*., 1999; Throup *et al*., 2000)
SP2022	STM240		*celB*	Cellobiose fTS System	
SP2162	STM265		*ptnC*	Mannose family PTS System	
SP2164	STM266		PM0834	Mannose family PTS system	
SP2236	STM281	*comD*		Histidine kinase	(Bartilson *et al*., 2001; Lau *et al*., 2001)
Transcription					
SP1156	STM147	*rhnB*		RNA degredation	
SP1483	STM178		*deaD2*	DEAD family-RNA Helicase	
Transport and binding proteins					
SP0078	STM12		*trkB*	Potassium uptake system	
SP0092	STM13		*ypcG*	ABC transporter, substrate binding protein	
SP0242	STM36		*bitB*	Iron ABC transporter, ATP-binding protein	
SP0479	STM68		*trkH*	Potassium uptake system	
SP0530	STM74, STM75	*blpA*		BlpC transport	

SP0600	STM79	*vex2*		Peptide transport	
SP0655	STM89		*ydiF*	Sodium:ion antiporter	
SP0720	STM97			ABC transporter, ATP-binding protein	
SP0729	STM100		*ctpA*	Cation transport	
SP1001	STM128		SPY0016	Amino acid permease	
SP1032	STM133, STM328	*pit2A*		Iron ABC transporter, Iron binding protein	(Brown *et al*., 2001)
SP1286	STM157	*uraA*		Uracil permease	
SP1321	STM160, STM161	*ntpK*		V-type Sodium ATPase, Subunit K	
SP1328	STM162		SA0303	Sodium:solute symporter	
SP1396	STM169		*pstB2*	Phosphate ABC transport, ATP-binding	
SP1398	STM170		*pstC*	Phosphate ABC transporter	
SP1399	STM171		*pstC2*	Phosphate ABC transporter	
SP1400	STM172		*pstS*	Phosphate ABC transporter	
SP1434	STM176		SA2216	ABC transporter, ATP-binding/permease	
SP1580	STM183	*msmK*		Multiple sugar transporter	(Polissi *et al*., 1998)
SP1715	STM192	*smtH*		ABC transporter, ATP-binding protein	(Lau *et al*., 2001)
SP1717	STM193		*ysdB*	ABC transporter, ATP binding	
SP1859	STM216			Unknown substrate	
SP1861	STM217		*proV*	Choline transporter	
SP1869	STM218		BH1205	Iron ABC transporter, permease protein	
SP1896	STM224	*msmF*		Multiple sugar transport	
SP1939	STM228	*dinF*		Damaged induced protein	
SP2086	STM245, STM246, STM247	*pstA*		Phosphate ABC transporter	
SP2101	STM250			Cation transport ATPase	
SP2108	STM252, STM253	*malX*		Maltose binding protein	
SP2170	STM268	*adcB*		Zinc transport	(Dintilhac *et al*., 1997)
SP2175	STM269, STM270		*dltB*	Lipidteichoic acid transport	
SP2231	STM279, STM280		SPY2211	ABC transporter, Permease	
Unknown function					
SP0049	STM3		*ykjJ*	vanZ, putative	
SP0121	STM20		SPY1846	metallo-β-lactamase family	
SP0267	STM43			Oxioreductase	
SP0320	STM53		*idnO*	Oxioreductase	
SP0571	STM76		XF1657	Cell-filamentation protein	
SP0622	STM81, STM82, STM83		SPY1069	NADH dehydrogenase	
SP0665	STM93		*pabB*	Chorismate binding protein	
SP0943	STM121	*gid*		Glucose inhibited divsion protein	
SP1023	STM131		SPY1144	GNAT family acetyltransferase	
SP1089	STM136		SPY1119	Glutamate amidotransferase	
SP1112	STM138		SPY1493	Unknown function	
SP1285	STM156	*gidB*		Glucose inhibited division protein	
SP1292	STM158, STM159			SAP domain protein	
SP1636	STM186		*ywmA*	Rrf2 familiy protein	
SP1646	STM188		SPY0646	metallo-β-lactamase family	
SP1941	STM229	*cinA*		Competence/damage induced protein	

[a] Indicates the gene name as assigned by the TIGR4 sequencing group

[b] The name of the gene that encodes the protein homologue, when available

[c] References are given only for genes which have been assigned a role in virulence or encode proteins with protective or immunogenic properties. References refer to the article that describes the role in virulence as denoted by TIGR (http://www.sciencemag.org/cgi/content/full/293/5529/498/DC1)

Table 2. Competition analysis of virulence gene tissue specificity of selected mutants identified by STM

Strain	TIGR ORF	Gene disrupted	Pneumonia Model		Bacteremia Model		Nasopharyngeal Model	
			Geometric mean *in vivo* CI[a]	*in vitro* CI[b]	Geometric mean *in vivo* CI	*in vitro* CI	Geometric mean *in vivo* CI	*in vitro* CI
Class I								
STM1	SP0023	*radA*			0.76 (3)	1.02		
STM23	SP0141	*mutR*	**0.064 (3)**	0.88				
STM29	SP0156	*rr07*	0.14 (4)	0.76	1.7 (3)	7.93	0.40 (8)	5.74
STM90	SP0661	*zmpR*	**<0.017 (4)**	0.92	0.55 (7)	0.92		
STM108	SP0829				0.63 (7)	1.30		
STM135	SP1045	SA1714			0.82 (4)	0.97		
STM139	SP1115	*mutR*	**<0.0079 (4)**	1.91	0.95 (4)	0.76		
STM241	SP2039	*sapR*	**<0.028 (7)**	0.57	0.43 (3)	1.48	0.13 (5)	0.59
STM244	SP2076	*hexA*			1.6 (4)	0.95		
STM256	SP2142	SPY1596			2.3 (3)	0.67		
Class II								
STM4	SP0050	*purH*			**<0.0067 (4)**	0.54		
STM119	SP0927	*smrC*	**<0.0081 (5)**	0.34	**<0.024 (8)**	0.34	1.85 (4)	1.02
STM125	SP0979	*pepF*			**0.21 (4)**	1.83		
STM175	SP1433		**<0.044 (4)**	0.83	**0.26 (8)**	0.98		
STM210	SP1830	*phoU*	**<0.011 (4)**	2.66	**0.019 (3)**	0.46	<0.086 (7)	2.66
STM237	SP2002	SA1157	**0.3 (5)**	0.72	**<0.073 (3)**	1.18		
STM330	SP1433		**<0.0092 (8)**	2.47	**<0.054 (6)**	1.27	<0.067 (3)	2.06
Class III								
STM64	SP0461	*rlrA*[c]	**<0.30 (7)**	1.90	0.74(4)	1.69	**<0.071 (10)**	1.10
STM124	SP0978	*coiA*	**<0.11(4)**	0.55	4.4 (3)	0.97	**0.037 (6)**	0.55
STM154	SP1278	*pyrR*	**<0.0038 (4)**	0.97	0.58 (3)	0.97	**<0.0086 (4)**	0.97
STM206	SP1800	*dmgB*	**<0.023 (4)**	0.36	0.56 (3)	0.82	**<0.0093 (6)**	0.36
Class IV								
STM38	SP0247	SPY2053	**<0.08 (4)**	1.34	**0.25 (4)**	0.41	**<0.09 (7)**	1.34
STM185	SP1633	*rr01*			**0.24 (8)**	0.63	**<0.024(5)**	0.63
STM208	SP1816	*trpG*	**<0.0022 (3)**	1.54	**<0.0058 (4)**	1.17	**<0.0088 (5)**	0.62
STM328	SP1032	*pit2A*	**0.27 (5)**	0.85	**0.27 (6)**	0.78	**<0.011 (4)**	1.03
STM329	SP2136	*pcpA*	**<0.015 (8)**	1.09	**<0.07 (7)**	1.09	**<0.053 (12)**	1.03

[a] The *in vivo* CI for each individual animal was calculated as the ratio of mutant to wild-type divided by the input ratio of mutant to wild-type bacteria. The geometric mean of the CIs is shown, and the number of animals infected in each experiment is indicated in parentheses. For competitions in which no mutant bacteria were recovered from a particular animal the number 1 was substituted as the numerator in determining the in vivo ratio for that animal, and thus the *in vivo mean* CI is denoted as less than the calculated value. Each in vivo competition was tested for statistical significance by the Student two-tailed t-test. p-values < 0.05 were considered significant and the corresponding mean is shown in bold.

[b] The *in vitro CI* was calculated as the ratio of mutant to wild-type bacteria after 5 h of growth in THY broth adjusted by the input ratio of mutant to wild-type bacteria.

[c] This gene description is provided in the present work

Table 3. Relevant strains, plasmids and primers used in this study

Strain	Relevant Genotype or Phenotype	Source or Reference
E. coli		
DH5αλpir	F⁻ Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR1 supE44 thi-1 gyrA96 relA1* λ::*pir*	(Hanahan, 1983; Kolter *et al.*, 1978)
S. pneumoniae		
TIGR4	Wild-type Type 4 encapsulated strain	Ingeborg Aaberge
AC353	Spontaneous Sm^R derivative of TIGR4	This work
AC846	AC353 *cps4E::magellan2* Cm^R	This work
AC1213	AC353 *rlrA::magellan2* Cm^R	This work
AC1214	AC353 *srtD::magellan2* Cm^R	This work
AC1215	AC353 *rrgA::magellan5* Spc^R	This work
AC1216	AC353 *rrgB::magellan5* Spc^R	This work
AC1217	AC353 *rrgC::magellan5* Spc^R	This work
AC1218	AC353 *srtB::magellan5* Spc^R	This work
AC1219	AC353 *srtC::magellan5* Spc^R	This work
Plasmids		
pEMCat	Contains *magellan2*; Ap^R, Cm^R	(Akerley *et al.*, 1998)
pEMSpc	Contains *magellan5*; Ap^R, Spc^R	(Martin *et al.*, 2000)
Primers (5' to 3')		
P6	GCAGATCTACCTACAACCTCAAGCT	This work
P7	CGAGATCTACCCATTCTAACCAAGC	This work
ARB1	GGCCACGCGTGCACTAGTAC$(N)_{10}$TACNG	(Merrell *et al.*,)
ARB2	GGCCACGCGTGCACTAGTAC	(Merrell *et al.*,)
MAG2F3	GGAATCATTTGAAGGTTGGTA	This work
MAG2F4	ACTAGCGACGCCATCTATGTG	This work
CPSF1	GCAGATAGTAAAAATAAAGGTGTAGAC	This work
CPSR1	TGCACTGAAGCCGAAGGCGACAAATGC	This work
TNPAB-F	GCCTCTTCCTGAGATTATGTCCTG	This work
REG2-R	ATTGCCGGTGTTATGTTCGTTTGG	This work
REG2-F	ATTTGTCCAAACGAACATAACACC	This work
PFL-R	TGAAAAATCTCTTGACTGGTTGAC	This work
marOUT	CCGGGGACTTATCAGCCAACC	This work

Table 4. Strains and plasmids used in this study

	Relevant genotype or sequence	Reference
Strains		
E. coli		
DH5αλpir	F⁻ Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR1 supE44 thi-1* gyrA96 *relA1 λ::pir*	(9, 12)
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacI^{q}ZDM15 Tn10 (Tet^r)]*	Stratagene
AC1287	DH5αλpir, contains pAC1287	This work
AC1288	DH5αλpir, contains pAC1288	This work
AC1289	DH5αλpir, contains pAC1289	This work
AC1290	DH5αλpir, contains pAC1290	This work
AC1291	DH5αλpir, contains pAC1291	This work
AC1292	XL-1 Blue, contains pAC1292	This work
S. pneumoniae		
AC353	TIGR4 Sm^r derivative	(10)
AC1213	*rlrA'::magellan2*, Sm^r, Cm^r	(10)
AC1278	*malM::rlrA::cat::malP*, Sm^r, Cm^r	This work
Plasmids		
pGEM-T	Cloning vector, Ap^r	Promega
pCR-Script Amp SK(+)	Cloning vector, Ap^r	Stratagene
pQE60	His_6 expression vector, Ap^r	Qiagen
pAC1000	*S. pneumoniae* suicide vector, Ap^r	This work
pCH84	pAC1000 *'malM::rlrA::cat::malP'*, Sm^r, Cm^r	This work
pAC1279	pGEM-T *rlrA* RPA probe, Ap^r	This work
pAC1280	pGEM-T *rrgA* RPA probe, Ap^r	This work
pAC1281	pGEM-T *rrgB* RPA probe, Ap^r	This work
pAC1282	pGEM-T *rrgC* RPA probe, Ap^r	This work
pAC1283	pGEM-T *srtB* RPA probe, Ap^r	This work
pAC1284	pGEM-T *srtC* RPA probe, Ap^r	This work
pAC1285	pGEM-T *srtD* RPA probe, Ap^r	This work
pAC1286	pGEM-T *rpoB* RPA probe, Ap^r	This work
pAC1287	pGEM-T *rlrA-rrgA* promoter fragment	This work
pAC1288	pGEM-T *rrgB* promoter fragment	This work
pAC1289	pGEM-T *rrgC* promoter fragment	This work
pAC1290	pGEM-T *srtB* promoter fragment	This work
pAC1291	pGEM-T *srtC-srtD* promoter fragment	This work
pAC1292	pQE60 *rlrA*-His_6, Ap^r	This work
pAC1293	pGEM-T *srtA* RPA probe	This work

Table 5. Sequences of the primers used in this study

Primer	Sequence (5' to 3')	SEQ ID NO:
MALFX	CCCTCGAGTGAAAGCTATCGTGAGCAATT	475
MALRP	CCGAGCTCAAGATCTGGATCCTTATTTCTTTAAATCTACC	476
MALPF2	CCCTCTAGAGAGCATGCGACAATAATCAGGAGACAAC	477
MALPRP	CCGCGGCTCGAGTTCAAGAGGCCATTTTTCAAG	478
PCATF1	CCCGGTCTAGAGTCGACGGTATCGATAAGCT	479
PCATR1	CCGGCGCATGCTTATAAAAGCCAGTCATTAG	480
RLRAFR	CGCGGATCCAAAGGAGAATCATCATGCTAAACAAATACATTGA	481
RLRARX	CCCTCTAGATTATAACAAATAGTGAGCCTT	482
PEVPF1	GAGGATCCTATACCGCGGCCATGTCTGCCCGTATT	483
PEVPR1	TTCACCACCTTTTCCCTAT	484
RLRAF2	TTACATGCTGTTTTATCAATAA	485
RLRAR7	AGTAGAAAGAAGCGGAGTATT	486
RRGAF3	CACTTTTATACGCTTTTGCTA	487
RRGAR3	TAATACGACTCACTATAGGTGCCATCCGTATTGTTTTTC	488
RRGBF2	AAACTATCATTGAAAGGGGAG	489
RRGBR1	TAATACGACTCACTATAGGGGCATTGCCCTGAGAGTTTA	490
RRGCF2	GGCTGCGATTATGGGTATT	491
RRGCR2	TAATACGACTCACTATAGGGGTCATCTCAAACGAAGTCT	492
SRTBF2	AGGACTGGGATTCTGATTTA	493
SRTBR1	TAATACGACTCACTATAGGATCGCCACTCACTACATTATT	494
SRTCF2	GATTCTTTTATGGATTATTCG	495
SRTCR2	TAATACGACTCACTATAGGGACGCCTTTCTTTTTCTCTTG	496
SRTDF2	GCGGTCATCCTTCTCTTGCT	497
SRTDR2	TAATACGACTCACTATAGGGTCGTCAGACACTTGGTAAT	498
SRTAF1	AAAAGAAAAACAAGCGAAAAA	499
SRTAR1	TCCTTCTCCCATTACTTGCTC	500
RPOBF3	TGCTTATGACTTGGCAGCAG	501
RPOBR3	GGCTTTCAATGCTTTCAATC	502
RLRAPE2	AGTTAAAGTAGACAGTTCATC	503
RRGAP2	ACGGATTACTTATGTTCTGAT	504
RRGBPE	GCTGAAAACAGGCTACTCGCT	505
RRGCPE	CCATAACAAAGAAGATACGACTAAT	506
SRTBPB	TTTTAAATCAGAATCCCAGTC	507
SRTCPE	GCGAATCCTACTAAGAAAATC	508
SRTDPE	TATCCCAATAAGGCTCGTAG	509
RLRA2	TGTGTGACCCAATCCATACTT	510
RRGA2	CCCTGTTTGTGGATACTGGTC	511
RRGB2	GGGTTACGAGTTTACGAATGA	512
RRGC2	CAATTGACTAACCACCTCCTG	513
SRTBP1	TCAGCAGTACCAGCATAAACC	514
SRTBP2	TTAAAAATAACAAGCGACCAC	515
SRTCD1	CCAAAACAATAAATAGGAATC	516
SRTCD2	CAAGTGGATCAAGTAAAGGTG	517
RLRAC1	CCATGGTTCTAAACAAATACATTGAAAAAA	518
RLRAC2	AGATCTTAACAAATAGTGAGCCTTTTTA	519

REGF1	TCTAGACATGTGTGTCTCCCTGTT	520
IIR1	TCTAGACATAGTTACCGAATCTTAGTT	521
AP2	AACAACTTCCATCACAATAGA	522
AP3	AGGATAGTTAATAGTAATACTATAC	523
AP4	TAACTATCCTAGTATAAATTAAAAC	524
AP5	TAAAACTCCACCAATACTCAT	525
AP6	ATGAGTATTGGTGGAGTTTTA	526

Table 6:

TIGR DESIGNATION	DNA SEQ ID NO:	AMINO ACID SEQ ID NO:	SEQUENCE CLAIMED IN PRESENT INVENTION?
SP0023	1	238	YES
SP0045	2	239	YES
SP0049	3	240	YES
SP0050	4	241	YES
SP0057	5	242	YES
SP0063	6	243	YES
SP0071	7	244	YES
SP0078	8	245	NO
SP0092	9	246	NO
SP0095	10	247	YES
SP0100	11	248	YES
SP0102	12	249	YES
SP0110	13	250	YES
SP0117	14	251	NO
SP0121	15	252	YES
SP0128	16	253	NO
SP0136	17	254	YES
SP0141	18	255	YES
SP0145	19	256	YES
SP0146	20	257	YES
SP0150	21	258	YES
SP0156	22	259	YES
SP0157	23	260	YES
SP0160	24	261	YES
SP0177	25	262	YES
SP0198	26	263	YES
SP0199	27	264	YES
SP0240	28	265	YES
SP0242	29	266	YES
SP0246	30	267	YES
SP0247	31	268	YES
SP0251	32	269	YES
SP0253	33	270	YES
SP0254	34	271	YES
SP0265	35	272	YES
SP0267	36	273	YES
SP0268	37	274	NO
SP0274	38	275	YES
SP0298	39	276	YES
SP0306	40	277	YES
SP0312	41	278	YES

SP0314	42	279	YES
SP0320	43	280	YES
SP0332	44	281	YES
SP0338	45	282	YES
SP0377	46	283	NO
SP0385	47	284	YES
SP0396	48	285	YES
SP0445	49	286	YES
SP0454	50	287	NO
SP0461	51	288	YES
SP0462	52	289	NO
SP0463	53	290	NO
SP0464	54	291	NO
SP0466	55	292	YES
SP0467	56	293	YES
SP0468	57	294	NO
SP0474	58	295	YES
SP0478	59	296	YES
SP0479	60	297	YES
SP0492	61	298	YES
SP0494	62	299	YES
SP0498	63	300	YES
SP0510	64	301	YES
SP0530	65	302	YES
SP0571	66	303	YES
SP0586	67	304	YES
SP0595	68	305	YES
SP0600	69	306	YES
SP0614	70	307	YES
SP0622	71	308	YES
SP0633	72	309	YES
SP0641	73	310	NO
SP0645	74	311	YES
SP0648	75	312	NO
SP0655	76	313	YES
SP0661	77	314	YES
SP0663	78	315	YES
SP0664	79	316	NO
SP0665	80	317	YES
SP0686	81	318	YES
SP0690	82	319	YES
SP0719	83	320	NO
SP0720	84	321	YES
SP0726	85	322	NO
SP0728	86	323	YES
SP0729	87	324	YES
SP0766	88	325	YES

SP0767	89	326	NO
SP0774	90	327	YES
SP0785	91	328	YES
SP0789	92	329	YES
SP0797	93	330	YES
SP0807	94	331	YES
SP0829	95	332	NO
SP0842	96	333	YES
SP0856	97	334	YES
SP0877	98	335	NO
SP0886	99	336	YES
SP0887	100	337	YES
SP0892	101	338	YES
SP0916	102	339	YES
SP0927	103	340	YES
SP0939	104	341	YES
SP0943	105	342	YES
SP0966	106	343	NO
SP0978	107	344	YES
SP0979	108	345	YES
SP0986	109	346	YES
SP1001	110	347	YES
SP1003	111	348	NO
SP1018	112	349	YES
SP1023	113	350	YES
SP1029	114	351	YES
SP1032	115	352	NO
SP1040	116	353	YES
SP1045	117	354	YES
SP1089	118	355	YES
SP1111	119	356	YES
SP1112	120	357	YES
SP1115	121	358	YES
SP1118	122	359	YES
SP1121	123	360	NO
SP1127	124	361	YES
SP1143	125	362	YES
SP1153	126	363	YES
SP1154	127	364	NO
SP1156	128	365	YES
SP1174	129	366	NO
SP1175	130	367	NO
SP1193	131	368	YES
SP1202	132	369	YES
SP1278	133	370	YES
SP1281	134	371	YES
SP1285	135	372	NO

SP1286	136	373	YES
SP1292	137	374	YES
SP1321	138	375	NO
SP1328	139	376	NO
SP1343	140	377	YES
SP1344	141	378	YES
SP1377	142	379	YES
SP1378	143	380	YES
SP1382	144	381	YES
SP1396	145	382	YES
SP1398	146	383	YES
SP1399	147	384	YES
SP1400	148	385	YES
SP1405	149	386	YES
SP1431	150	387	YES
SP1433	151	388	YES
SP1434	152	389	YES
SP1483	153	390	YES
SP1518	154	391	YES
SP1529	155	392	YES
SP1538	156	393	YES
SP1544	157	394	YES
SP1580	158	395	YES
SP1591	159	396	YES
SP1633	160	397	YES
SP1636	161	398	YES
SP1645	162	399	YES
SP1646	163	400	YES
SP1652	164	401	YES
SP1654	165	402	YES
SP1706	166	403	NO
SP1715	167	404	YES
SP1717	168	405	YES
SP1760	169	406	YES
SP1770	170	407	YES
SP1771	171	408	NO
SP1772	172	409	NO
SP1779	173	410	YES
SP1780	174	411	YES
SP1793	175	412	YES
SP1800	176	413	YES
SP1815	177	414	YES
SP1816	178	415	YES
SP1817	179	416	YES
SP1830	180	417	YES
SP1847	181	418	YES
SP1854	182	419	YES

SP1855	183	420	YES
SP1856	184	421	YES
SP1859	185	422	NO
SP1861	186	423	NO
SP1869	187	424	YES
SP1879	188	425	YES
SP1889	189	426	YES
SP1890	190	427	YES
SP1891	191	428	YES
SP1896	192	429	NO
SP1898	193	430	YES
SP1923	194	431	YES
SP1939	195	432	YES
SP1941	196	433	YES
SP1952	197	434	YES
SP1956	198	435	YES
SP1964	199	436	YES
SP1970	200	437	YES
SP1976	201	438	YES
SP1998	202	439	YES
SP2002	203	440	YES
SP2017	204	441	YES
SP2022	205	442	YES
SP2039	206	443	YES
SP2052	207	444	YES
SP2060	208	445	YES
SP2076	209	446	YES
SP2086	210	447	YES
SP2095	211	448	YES
SP2098	212	449	YES
SP2101	213	450	YES
SP2105	214	451	YES
SP2108	215	452	YES
SP2128	216	453	YES
SP2131	217	454	YES
SP2136	218	455	NO
SP2142	219	456	YES
SP2143	220	457	YES
SP2145	221	458	YES
SP2146	222	459	YES
SP2159	223	460	NO
SP2162	224	461	YES
SP2164	225	462	YES
SP2167	226	463	YES
SP2170	227	464	YES
SP2175	228	465	NO
SP2176	229	466	YES

SP2182	230	467	YES
SP2190	231	468	NO
SP2201	232	469	YES
SP2209	233	470	YES
SP2210	234	471	YES
SP2231	235	472	YES
SP2236	236	473	YES
SP2239	237	474	YES

REFERENCES

The references cited herein are incorporated by reference.

1. Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T., Dormitzer, M., Dagan, R., Brewah, Y.A., Barren, P., Lathigra, R., Langermann, S., Koenig, S. and Johnson, S. (2001) Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis. *Infect Immun*, **69**, 949-958.

2. Akerley, B.J., Rubin, E.J., Camilli, A., Lampe, D.J., Robertson, H.M. and Mekalanos, J.J. (1998) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A*, **95**, 8927-8932.

3. Autret, N., Dubail, I., Trieu-Cuot, P., Berche, P. and Charbit, A. (2001) Identification of new genes involved in the virulence of Listeria monocytogenes by signature-tagged transposon mutagenesis. *Infect Immun*, **69**, 2054-2065.

4. Bartilson, M., Marra, A., Christine, J., Asundi, J.S., Schneider, W.P. and Hromockyj, A.E. (2001) Differential fluorescence induction reveals Streptococcus pneumoniae loci regulated by competence stimulatory peptide. *Mol Microbiol*, **39**, 126-135.

5. Berry, A.M., Lock, R.A., Thomas, S.M., Rajan, D.P., Hansman, D. and Paton, J.C. (1994) Cloning and nucleotide sequence of the Streptococcus pneumoniae hyaluronidase gene and purification of the enzyme from recombinant Escherichia coli. *Infect Immun*, **62**, 1101-1108.

6. Berry, A.M. and Paton, J.C. (2000) Additive attenuation of virulence of Streptococcus pneumoniae by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun*, **68**, 133-140.

7. Bongaerts, R.J., Heinz, H.P., Hadding, U. and Zysk, G. (2000) Antigenicity, expression, and molecular characterization of surface- located pullulanase of Streptococcus pneumoniae. *Infect Immun*, **68**, 7141-7143.

8. Bricker, A.L. and Camilli, A. (1999) Transformation of a type 4 encapsulated strain of Streptococcus pneumoniae. *FEMS Microbiol Lett*, **172**, 131-135.

9. Briles, D.E., Hollingshead, S., Brooks-Walter, A., Nabors, G.S., Ferguson, L., Schilling, M., Gravenstein, S., Braun, P., King, J. and Swift, A. (2000a) The potential to use PspA and other pneumococcal proteins to elicit protection against pneumococcal infection. *Vaccine*, **18**, 1707-1711.

10. Briles, D.E., Hollingshead, S.K., Nabors, G.S., Paton, J.C. and Brooks-Walter, A. (2000b) The potential for using protein vaccines to protect against otitis media caused by Streptococcus pneumoniae. *Vaccine*, **19**, S87-95.

11. Brown, J.S., Gilliland, S.M. and Holden, D.W. (2001) A Streptococcus pneumoniae pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol Microbiol*, **40**, 572-585.

12. Cundell, D.R., Pearce, B.J., Sandros, J., Naughton, A.M. and Masure, H.R. (1995) Peptide permeases from Streptococcus pneumoniae affect adherence to eucaryotic cells. *Infect Immun*, **63**, 2493-2498.

13. Dintilhac, A., Alloing, G., Granadel, C. and Claverys, J.P. (1997) Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol*, **25**, 727-739.

14. Fogg, G.C., Gibson, C.M. and Caparon, M.G. (1994) The identification of rofA, a positive-acting regulatory component of prtF expression: use of an m gamma delta-based shuttle mutagenesis strategy in Streptococcus pyogenes. *Mol Microbiol*, **11**, 671-684.

15. Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I. and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infect Immun*, **68**, 5690-5695.

16. Hammerschmidt, S., Bethe, G., Remane, P.H. and Chhatwal, G.S. (1999) Identification of pneumococcal surface protein A as a lactoferrin- binding protein of Streptococcus pneumoniae. *Infect Immun*, **67**, 1683-1687.

17. Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166**, 557-580.

18. Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E. and Holden, D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science*, **269**, 400-403.

19. Hollingshead, S.K., Becker, R. and Briles, D.E. (2000) Diversity of PspA: mosaic genes and evidence for past recombination in Streptococcus pneumoniae. *Infect Immun*, **68**, 5889-5900.

20. Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R., Jr., Skatrud, P.L. and Glass, J.I. (2001) Genome of the bacterium Streptococcus pneumoniae strain R6. *J Bacteriol*, **183**, 5709-5717.

21. Jedrzejas, M.J. (2001) Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev*, **65**, 187-207 ; first page, table of contents.

22. Jones, A.L., Knoll, K.M. and Rubens, C.E. (2000) Identification of Streptococcus agalactiae virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. *Mol Microbiol*, **37**, 1444-1455.

23. Kolter, R., Inuzuka, M. and Helinski, D.R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell*, **15**, 1199-1208.

24. Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J. and Robertson, H.M. (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci U S A*, **96**, 11428-11433.

25. Lampe, D.J., Churchill, M.E. and Robertson, H.M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *Embo J*, **15**, 5470-5479.

26. Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck, W. and Amrein, K.E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae. *Gene*, **237**, 223-234.

27. Lau, G.W., Haataja, S., Lonetto, M., Kensit, S.E., Marra, A., Bryant, A.P., McDevitt, D., Morrison, D.A. and Holden, D.W. (2001) A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. *Mol Microbiol*, **40**, 555-571.

28. Martin, B., Prudhomme, M., Alloing, G., Granadel, C. and Claverys, J.P. (2000) Cross-regulation of competence pheromone production and export in the early control of transformation in Streptococcus pneumoniae. *Mol Microbiol*, **38**, 867-878.

29. Mazmanian, S.K., Ton-That, H. and Schneewind, O. (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol Microbiol*, **40**, 1049-1057.

30. Mazmanian, S.K., Ton-That, H., Su, K. and Schneewind, O. (2002) An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis. *Proc Natl Acad Sci U S A*, **99**, 2293-2298.

31. McCullers, J.A. and Tuomanen, E.I. (2001) Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci*, **6**, D877-889.

32. Mei, J.M., Nourbakhsh, F., Ford, C.W. and Holden, D.W. (1997) Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol Microbiol*, **26**, 399-407.

33. Merrell, D.S., Hava, D.L. and Camilli, A. (2002) Identification of Novel Factors Involved in Colonization and Acid Tolerance of *Vibrio cholerae*. *(in press)*.

34. Pallen, M.J., Lam, A.C., Antonio, M. and Dunbar, K. (2001) An embarrassment of sortases - a richness of substrates? *Trends Microbiol*, **9**, 97-101.

35. Paton, J.C. and Giammarinaro, P. (2001) Genome-based analysis of pneumococcal virulence factors: the quest for novel vaccine antigens and drug targets. *Trends Microbiol*, **9**, 515-518.

36. Pestova, E.V., Havarstein, L.S. and Morrison, D.A. (1996) Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol*, **21**, 853-862.

37. Podbielski, A., Woischnik, M., Leonard, B.A. and Schmidt, K.H. (1999) Characterization of nra, a global negative regulator gene in group A streptococci. *Mol Microbiol*, **31**, 1051-1064.

38. Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L. and Simon, D. (1998) Large-scale identification of virulence genes from Streptococcus pneumoniae. *Infect Immun*, **66**, 5620-5629.

39. Poulsen, K., Reinholdt, J., Jespersgaard, C., Boye, K., Brown, T.A., Hauge, M. and Kilian, M. (1998) A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: evidence for recombination within and between species. *Infect Immun*, **66**, 181-190.

40. Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H.R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. *Mol Microbiol*, **25**, 819-829.

41. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1998) *Molecular Cloning: A Laboratory Manual, Second Edition.* Cold Spring Harbor Lab. Press, Plainview, NY.

42. Sanchez-Beato, A.R., Lopez, R. and Garcia, J.L. (1998) Molecular characterization of PcpA: a novel choline-binding protein of Streptococcus pneumoniae. *FEMS Microbiol Lett*, **164**, 207-214.

43. Schneewind, O., Mihaylova-Petkov, D. and Model, P. (1993) Cell wall sorting signals in surface proteins of gram-positive bacteria. *Embo J*, **12**, 4803-4811.

44. Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O.,

Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science*, **293**, 498-506.

45. Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M. and Burnham, M.K. (2000) A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. *Mol Microbiol*, **35**, 566-576.

46. Tomasz, A. (1999) New faces of an old pathogen: emergence and spread of multidrug- resistant Streptococcus pneumoniae. *Am J Med*, **107**, 55S-62S.

47. Unsworth, K.E. and Holden, D.W. (2000) Identification and analysis of bacterial virulence genes in vivo. *Philos Trans R Soc Lond B Biol Sci*, **355**, 613-622.

48. Walker, J.A., Allen, R.L., Falmagne, P., Johnson, M.K. and Boulnois, G.J. (1987) Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of Streptococcus pneumoniae. *Infect Immun*, **55**, 1184-1189.

49. Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S. and Koenig, S. (2001) Use of a whole genome

approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. *Infect Immun*, **69**, 1593-1598.

50. Wu, H.Y., Virolainen, A., Mathews, B., King, J., Russell, M.W. and Briles, D.E. (1997) Establishment of a Streptococcus pneumoniae nasopharyngeal colonization model in adult mice. *Microb Pathog*, **23**, 127-137.

51. Yesilkaya, H., Kadioglu, A., Gingles, N., Alexander, J.E., Mitchell, T.J. and Andrew, P.W. (2000) Role of manganese-containing superoxide dismutase in oxidative stress and virulence of Streptococcus pneumoniae. *Infect Immun*, **68**, 2819-2826.

52. Yother, J. and White, J.M. (1994) Novel surface attachment mechanism of the Streptococcus pneumoniae protein PspA. *J Bacteriol*, **176**, 2976-2985.

53. Zysk, G., Bongaerts, R.J., ten Thoren, E., Bethe, G., Hakenbeck, R. and Heinz, H.P. (2000) Detection of 23 immunogenic pneumococcal proteins using convalescent- phase serum. *Infect Immun*, **68**, 3740-3743.

54. Acebo, P., C. Nieto, M. A. Corrales, M. Espinosa, and P. Lopez. 2000. Quantitative detection of *Streptococcus pneumoniae* cells harbouring single or multiple copies of the gene encoding the green fluorescent protein. Microbiology 146:1267-1273.

55. Charpentier, E., and E. Tuomanen. 2000. Mechanisms of antibiotic resistance and tolerance in *Streptococcus pneumoniae*. Microbes Infect 2:1855-1864.

56. Claverys, J. P., A. Dintilhac, E. V. Pestova, B. Martin, and D. A. Morrison. 1995. Construction and evaluation of new drug-resistance cassettes for gene

disruption mutagenesis in *Streptococcus pneumoniae*, using an ami test platform. Gene 164:123-128.

57. Crook, D. W., and B. G. Spratt. 1998. Multiple antibiotic resistance in *Streptococcus pneumoniae*. Br Med Bull 54:595-610.

58. Fogg, G. C., and M. G. Caparon. 1997. Constitutive expression of fibronectin binding in *Streptococcus pyogenes* as a result of anaerobic activation of *rofA*. J Bacteriol 179:6172-6180.

59. Fogg, G. C., C. M. Gibson, and M. G. Caparon. 1994. The identification of *rofA*, a positive-acting regulatory component of *prtF* expression: use of an mγδ-based shuttle mutagenesis strategy in *Streptococcus pyogenes*. Mol Microbiol 11:671-684.

60. Gillespie, S. H., and I. Balakrishnan. 2000. Pathogenesis of pneumococcal infection. J Med Microbiol 49:1057-1067.

61. Granok, A. B., D. Parsonage, R. P. Ross, and M. G. Caparon. 2000. The RofA binding site in *Streptococcus pyogenes* is utilized in multiple transcriptional pathways. J Bacteriol 182:1529-1540.

62. Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166:557-580.

63. Hava, D. L., and A. Camilli. 2002. Identification of tissue-specific pneumococcal virulence factors. Molecular Microbiology (in press).

64. Hoskins, J., W. E. Alborn, Jr., J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J.

S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, Jr., P. L. Skatrud, and J. I. Glass. 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol 183:5709-5717.

65. Kolter, R., M. Inuzuka, and D. R. Helinski. 1978. Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. Cell 15:1199-1208.

66. Lau, G. W., S. Haataja, M. Lonetto, S. E. Kensit, A. Marra, A. P. Bryant, D. McDevitt, D. A. Morrison, and D. W. Holden. 2001. A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. Mol Microbiol 40:555-571.

67. Lee, M. S., and D. A. Morrison. 1999. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. J Bacteriol 181:5004-5016.

68. Mazmanian, S. K., G. Liu, H. Ton-That, and O. Schneewind. 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science 285:760-763.

69. Mazmanian, S. K., H. Ton-That, and O. Schneewind. 2001. Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Mol Microbiol 40:1049-1057.

70. Mazmanian, S. K., H. Ton-That, K. Su, and O. Schneewind. 2002. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. Proc Natl Acad Sci U S A 99:2293-2298.

71. McCullers, J. A., and E. I. Tuomanen. 2001. Molecular pathogenesis of pneumococcal pneumonia. Front Biosci 6:D877-889.

72. Merrell, D. S., and A. Camilli. 2000. Regulation of *Vibrio cholerae* genes required for acid tolerance by a member of the "ToxR-like" family of transcriptional regulators. J Bacteriol 182:5342-5350.

73. Pallen, M. J., A. C. Lam, M. Antonio, and K. Dunbar. 2001. An embarrassment of sortases - a richness of substrates? Trends Microbiol 9:97-101.

74. Pestova, E. V., and D. A. Morrison. 1998. Isolation and characterization of three *Streptococcus pneumoniae* transformation-specific loci by use of a lacZ reporter insertion vector. J Bacteriol 180:2701-2710.

75. Podbielski, A., M. Woischnik, B. A. Leonard, and K. H. Schmidt. 1999. Characterization of nra, a global negative regulator gene in group A streptococci. Mol Microbiol 31:1051-1064.

76. Polissi, A., A. Pontiggia, G. Feger, M. Altieri, H. Mottl, L. Ferrari, and D. Simon. 1998. Large-scale identification of virulence genes from *Streptococcus pneumoniae*. Infect Immun 66:5620-5629.

77. Puyet, A., and M. Espinosa. 1993. Structure of the maltodextrin-uptake locus of *Streptococcus pneumoniae*. Correlation to the Escherichia coli maltose regulon. J Mol Biol 230:800-811.

78. Sabelnikov, A. G., B. Greenberg, and S. A. Lacks. 1995. An extended -10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. J Mol Biol 250:144-155.

79. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293:498-506.

80. Tomasz, A. 1997. Antibiotic resistance in *Streptococcus pneumoniae*. Clin Infect Dis 24:S85-88.

81. Tomasz, A. 1999. New faces of an old pathogen: emergence and spread of multidrug-resistant *Streptococcus*. AM J Med. 107(1A):55S-62S.

82. Unsworth KE, Holden DW. 2000. Identification and analysis of bacterial virulence genes *in vivo*. Philos Trans R Soc Lond B Biol Sci. 355(1397):613-22.

All references cited herein and throughout the specification are hereby incorporated by reference in their entirety.

We claim:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of: (a) an isolated nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-7, 10-13, 15, 17-36, 38-45, 47-49, 51, 55-56, 58-72, 74, 76-78, 80-82, 84, 86-88, 90-94, 96-97, 99-105, 107-110, 112-114, 116-122, 124-126, 128, 131-134, 136-137, 140-165, 167-170, 173-184, 187-191, 193-217, 219-222, 224-227, 229-230, and 232-237; (b) a nucleotide sequence encoding an amino acid sequence set forth in a sequence selected from the group consisting of SEQ ID NOs.: 238-244, 247-250, 252, 254-273, 275-282, 284-286, 288, 292-293, 295-309, 311, 313-315, 317-319, 321, 323-325, 327-331, 333-334, 336-342, 344-347, 349-351, 353-359, 361-363, 365, 368-371, 373-374, 377-402, 404-407, 410-421, 424-428, 430-454, 456-459, 461-464, 466-467, and 469-474; (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a); or (d) a nucleotide sequence complementary to any of the nucleotide sequences in (b).

2. An isolated unique fragment of the isolated nucleotide sequence of claim 1.

3. An isolated nucleotide probe comprising a nucleotide sequence that is capable of hybridizing to the isolated nucleotide sequence of claim 1.

4. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a polypeptide having an amino acid sequence in (b) of claim 1.

5. An isolated recombinant vector comprising the nucleotide sequence of claim 1, or a fragment thereof.

6. A recombinant host cell transformed with the vector of claim 5.

7. A method of producing a polypeptide comprising culturing the host cell of claim 6 under conditions favoring expressing the nucleotide sequence..

8. An isolated polypeptide encoded by a nucleic acid molecule comprising a sequence selected from the group consisting of: (a) an isolated nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-7, 10-13, 15, 17-36, 38-45, 47-49, 51, 55-56, 58-72, 74, 76-78, 80-82, 84, 86-88, 90-94, 96-97, 99-105, 107-110, 112-114, 116-122, 124-126, 128, 131-134, 136-137, 140-165, 167-170, 173-184, 187-191, 193-217, 219-222, 224-227, 229-230, and 232-237; (b) a nucleotide sequence encoding an amino acid sequence set forth in a sequence selected from the group consisting of SEQ ID NOs.: 238-244, 247-250, 252, 254-273, 275-282, 284-286, 288, 292-293, 295-309, 311, 313-315, 317-319, 321, 323-325, 327-331, 333-334, 336-342, 344-347, 349-351, 353-359, 361-363, 365, 368-371, 373-374, 377-402, 404-407, 410-421, 424-428, 430-454, 456-459, 461-464, 466-467, and 469-474; (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a); or (d) a nucleotide sequence complementary to any of the nucleotide sequences in (b).

9. A unique fragment of the isolated polypeptide of claim 8.

10. An isolated polypeptide antigen comprising an amino acid sequence, or a fragment thereof, of the isolated polypeptide of claim 8.

11. An isolated nucleic acid molecule comprising a polynucleotide with a nucleotide sequence encoding the polypeptide of claim 8.

12. An isolated antibody that binds specifically to a polypeptide of claim 8.

13. A vaccine, comprising: at least one *Streptococcus pneumoniae* polypeptide of claim 8 or at least one nucleic acid molecule of claim 1, and a pharmaceutically acceptable diluent, carrier, or excipient; wherein said polypeptide is present in an amount effective to elicit protective antibodies in an animal to a member of the *Streptococcus* genus.

14. The vaccine composition of claim 13 further comprising an adjuvant.

15. A method of preventing or attenuating an infection caused by a member of the *Streptococcus* genus in an animal, comprising administering to said animal the vaccine of claim 13.

16. A method of detecting *Streptococcus* nucleic acids in a biological sample obtained from an animal comprising: (a) contacting the sample with one or more of the probes of claim 3, under conditions such that hybridization occurs, and (b) detecting hybridization of said one or more probes to the one or more *Streptococcus* nucleic acid sequences present in the biological sample.

17. A method of detecting *Streptococcus* nucleic acids in a biological sample obtained from an animal, comprising: (a) amplifying one or more *Streptococcus* nucleic acid sequences in said sample using polymerase chain reaction, (b) contacting the amplified *Streptococcus* nucleic acid sequence(s) with one or more of the probes of claim 3, under conditions such that hybridization occurs, and (c) detecting hybridization of said one or more probes to the one or more of the amplified *Streptococcus* nucleic acid sequences.

18. A kit for detecting *Streptococcus* antibodies in a biological sample obtained from an animal, comprising (a) a polypeptide of claim 8 attached to a solid support; and (b) detecting means.

19. A method of detecting *Streptococcus* antibodies in a biological sample obtained from an animal, comprising (a) contacting the sample with a polypeptide of claim 8; and (b) detecting antibody-antigen complexes.

20. A pharmaceutical composition for reducing the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy and/or for treating *Streptococcus pneumoniae* infections

comprising the antibody of claim 12 and a pharmaceutically acceptable carrier.

21. A method for the treatment of *Streptococcus pneumoniae* infections comprising administering to an individual in need thereof a therapeutically effective amount of the pharmaceutical composition of claim 20 to treat *Streptococcus pneumoniae* infection.

22. A method for reducing the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy, comprising administering to a population of individuals a pharmaceutical composition according to claim 20, to reduce the occurrence of *Streptococcus pneumoniae* infections in the population.

23. A method for the treatment of *Streptococcus pneumoniae* infections comprising administering to an individual in need a therapeutically effective amount of the antibody of claim 12 to treat *Streptococcus pneumoniae* infection.

24. A method for reducing the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy, comprising administering to a population of individuals an antibody of claim 12, to reduce the occurrence of *Streptococcus pneumoniae* infections in the population.

25. A pharmaceutical composition for reducing the occurrence of *Streptococcus pneumoniae* infections in a population of individuals by passive immunotherapy, and/or for treating *Streptococcus pneumoniae* infections comprising as an active ingredient at least one antibody in accordance with claim 12 in combination with at least one other active ingredient being an anti viral agent.

26. A method for the diagnosis of *Streptococcus pneumoniae* infections in a body fluid sample comprising: (a) contacting said sample with an antibody of claim

12 under conditions enabling the formation of antibody-antigen complexes; (b) determining the level of antibody-antigen complexes formed, wherein a determination of the presence of a level of antibody-antigen complexes significantly higher than that formed in a control sample indicates a *Streptococcus pneumoniae* infection in the tested body fluid sample.

27. A method for the identification of an agent that is effective in the treatment and/or diagnosis of *Streptococcus pneumoniae* infection, comprising contacting a polypeptide of SEQ ID NO: 238-474 with a target compound, and selecting a compound that binds specifically to said nucleic acid or polypeptide.

28. The use of the agent of claim 27 in the manufacture of a medicament for use in the treatment or prophylaxis of *Streptococcus pneumoniae* infection.

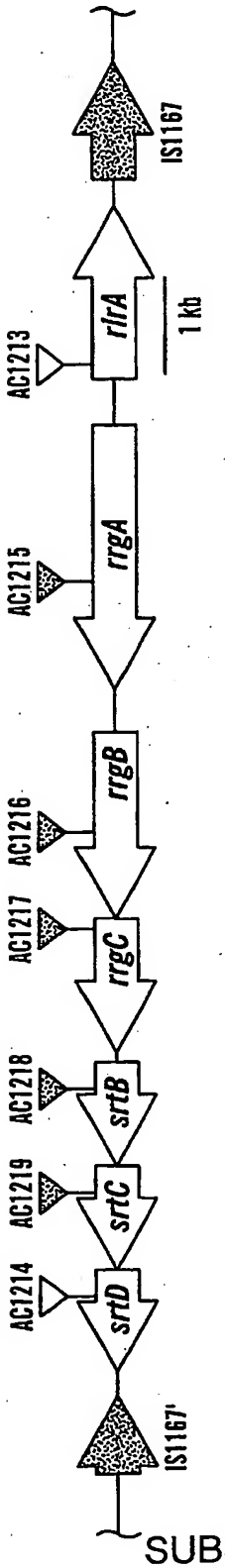


FIG. 1A

PROTEIN	C-TERMINAL SEQUENCE
RrgA	YPRTGGIGMLPFYLIGCMMMGGVLLYTRKIIP
RrgB	IPQFGGIGTHFAVAGAAIMGIAVYAYVKNNKDEDQLA
RrgC	VPDTGEETLYILMLVAILLFGSGYYLTKKPNN

FIG. 1B

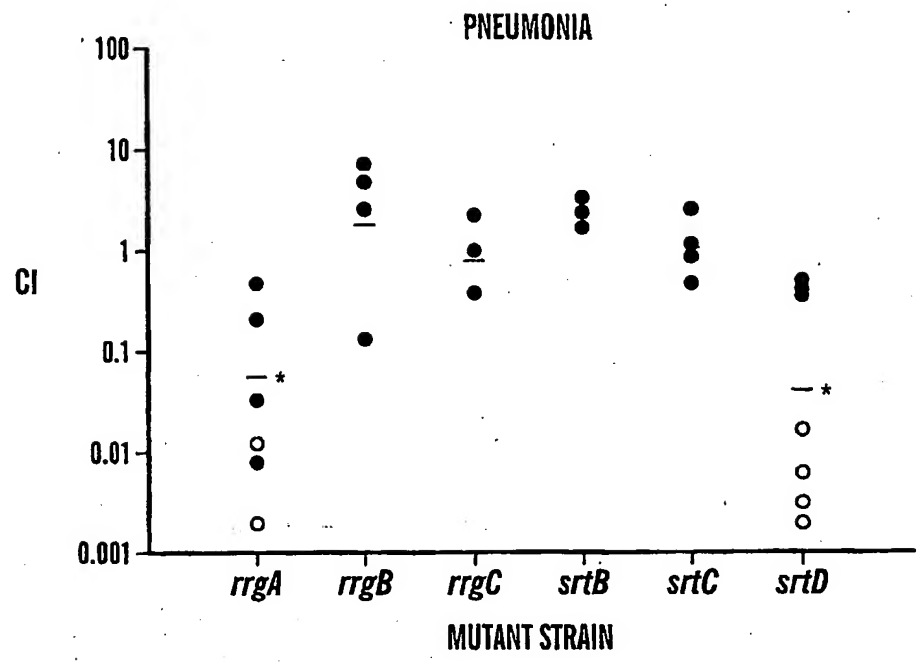

PNEUMONIA
CI
100
10
1
0.1
0.01
0.001
rrgA
rrgB
rrgC
srtB
srtC
srtD
MUTANT STRAIN

FIG. 2A

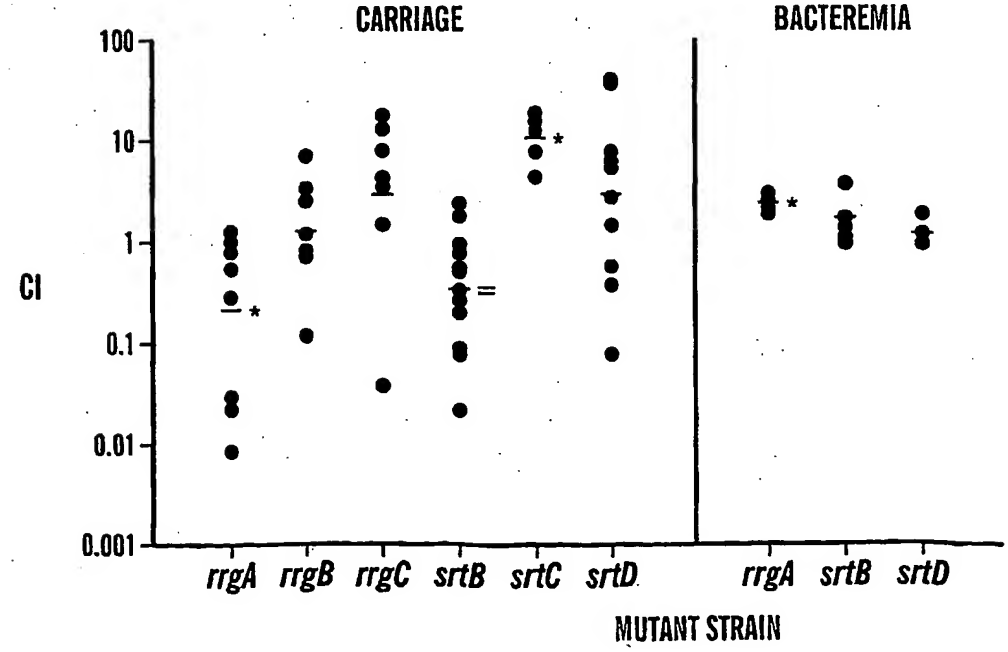

CARRIAGE
BACTEREMIA
CI
100
10
1
0.1
0.01
0.001
rrgA
rrgB
rrgC
srtB
srtC
srtD
rrgA
srtB
srtD
MUTANT STRAIN

FIG. 2B

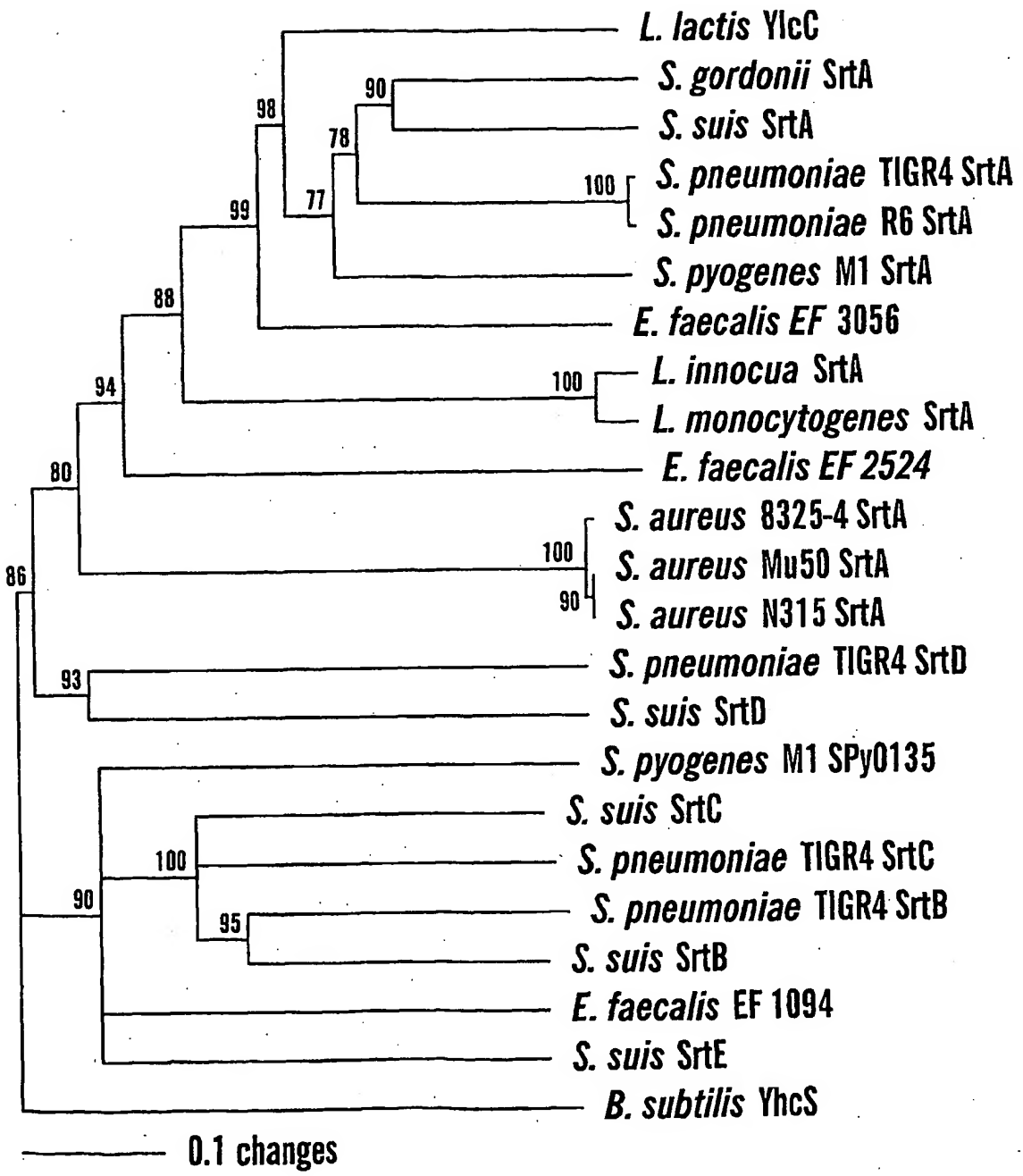

L. lactis YlcC
S. gordonii SrtA
S. suis SrtA
S. pneumoniae TIGR4 SrtA
S. pneumoniae R6 SrtA
S. pyogenes M1 SrtA
E. faecalis EF 3056
L. innocua SrtA
L. monocytogenes SrtA
E. faecalis EF 2524
S. aureus 8325-4 SrtA
S. aureus Mu50 SrtA
S. aureus N315 SrtA
S. pneumoniae TIGR4 SrtD
S. suis SrtD
S. pyogenes M1 SPy0135
S. suis SrtC
S. pneumoniae TIGR4 SrtC
S. pneumoniae TIGR4 SrtB
S. suis SrtB
E. faecalis EF 1094
S. suis SrtE
B. subtilis YhcS
0.1 changes
98
90
78
100
77
99
88
94
100
80
86
100
90
93
100
90
95

FIG. 3

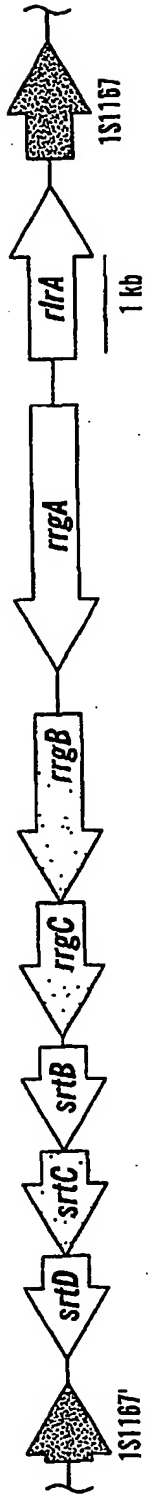


FIG. 4

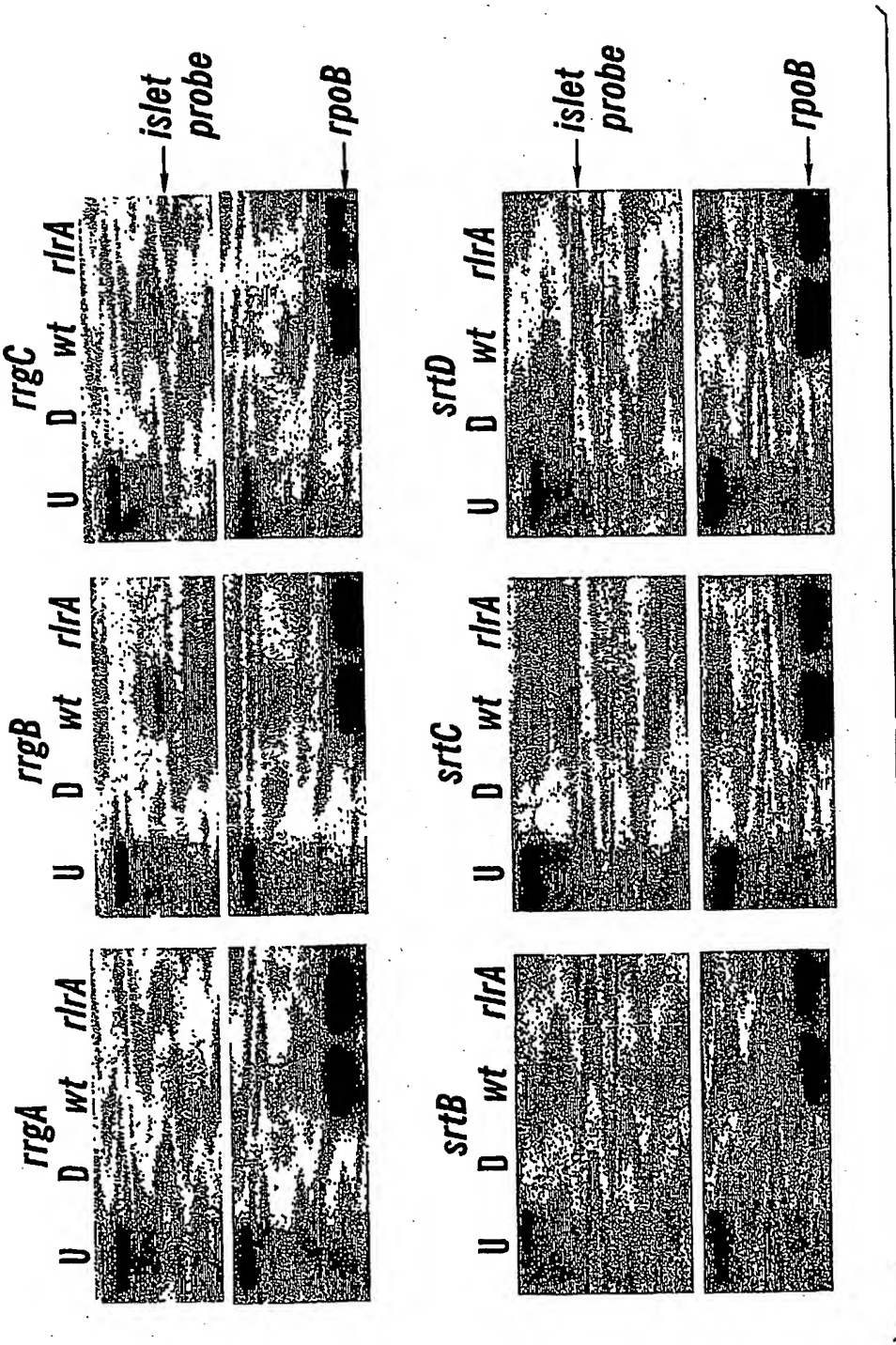

rrgA
U D wt rlrA
rrgB
U D wt rlrA
rrgC
U D wt rlrA
islet probe
rpoB
srtB
U D wt rlrA
srtC
U D wt rlrA
srtD
U D wt rlrA
islet probe
rpoB

FIG. 5A

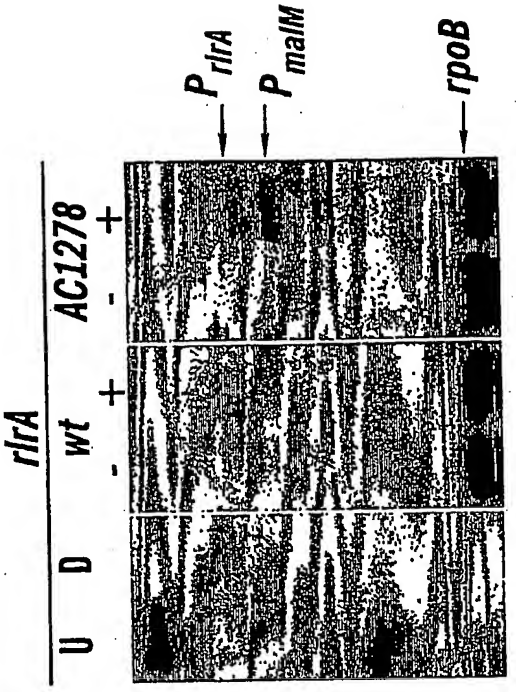


FIG. 5C

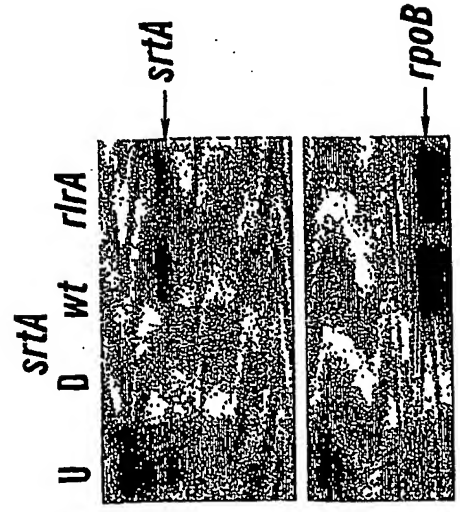


FIG. 5B

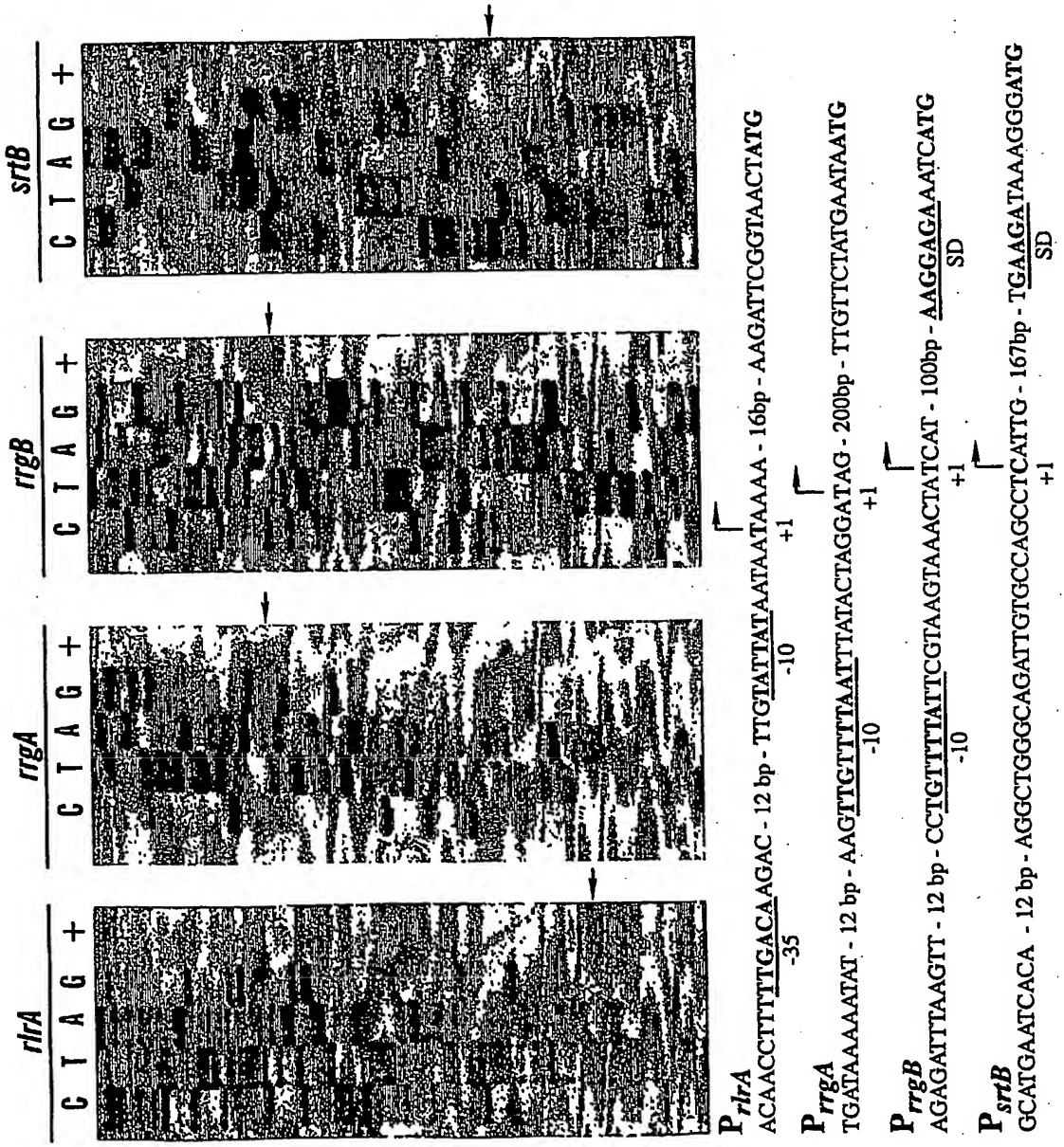

rlrA
C T A G +
rrgA
C T A G +
rrgB
C T A G +
srtB
C T A G +
P rlrA
ACAACCTTTTGACAAGAC - 12 bp - TTGTATTATAATAAATAAAA - 16bp - AAGATTCGGTAACTATG
-35 -10 +1
P rrgA
TGATAAAAATAT - 12 bp - AAGTTGTTTTAATTTATACTAGGATAG - 200bp - TTGTTCTATGAATAATG
-10 +1
P rrgB
AGAGATTTAAGTT - 12 bp - CCTGTTTTTATTCGTAAGTAAACTATCAT - 100bp - AAGGAGAAATCATG
-10 +1 SD
P srtB
GCATGAATCACA - 12 bp - AGGCTGGGCAGATTGTGCCAGCCTCATTG - 167bp - TGAAGATAAAGGGATG
+1 SD

FIG. 6

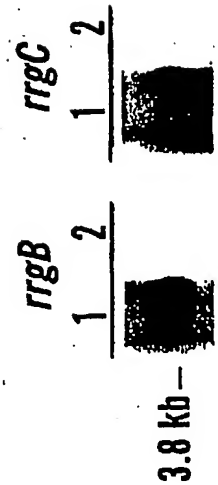


FIG. 7A

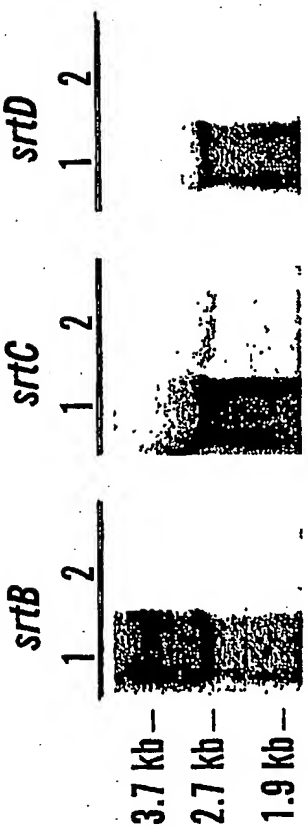


FIG. 7B

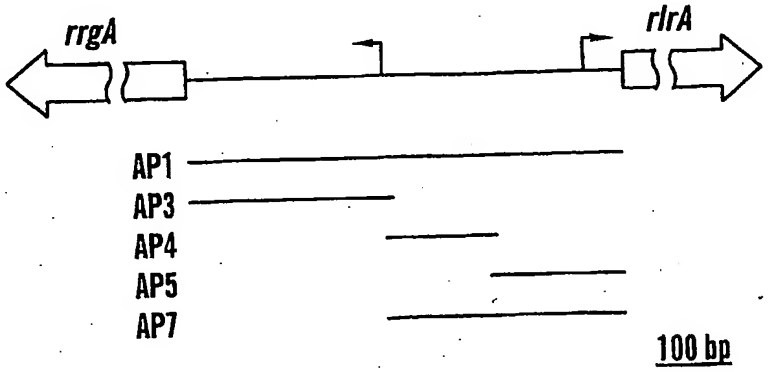

rrgA
rlrA
AP1
AP3
AP4
AP5
AP7
100 bp

FIG. 8A

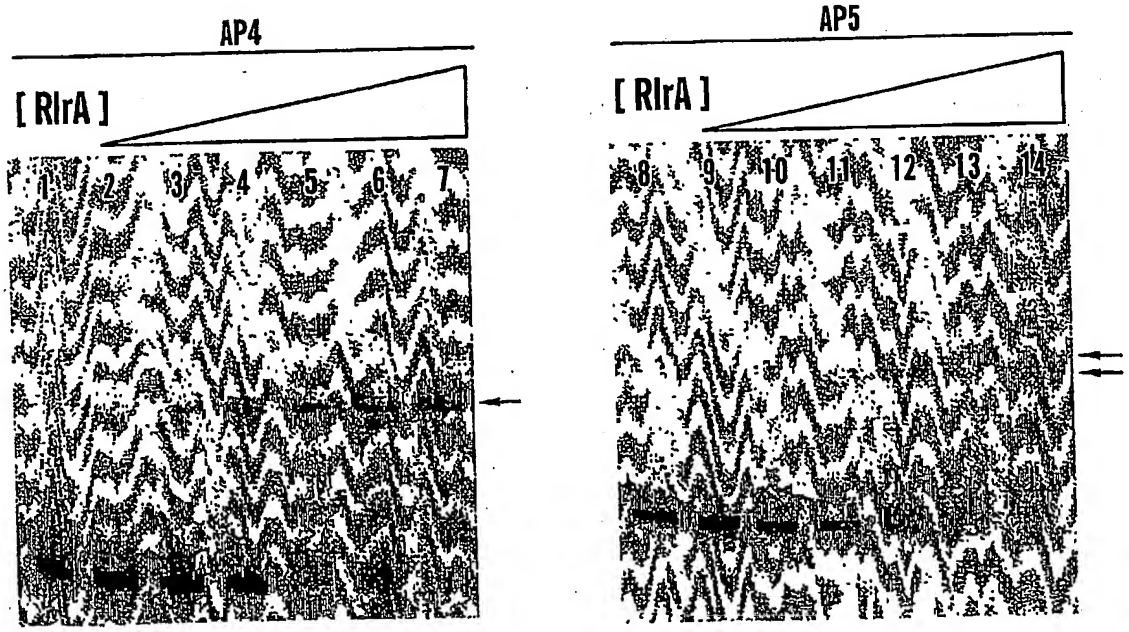

AP4
[RlrA]
1 2 3 4 5 6 7
AP5
[RlrA]
8 9 10 11 12 13 14

FIG. 8B

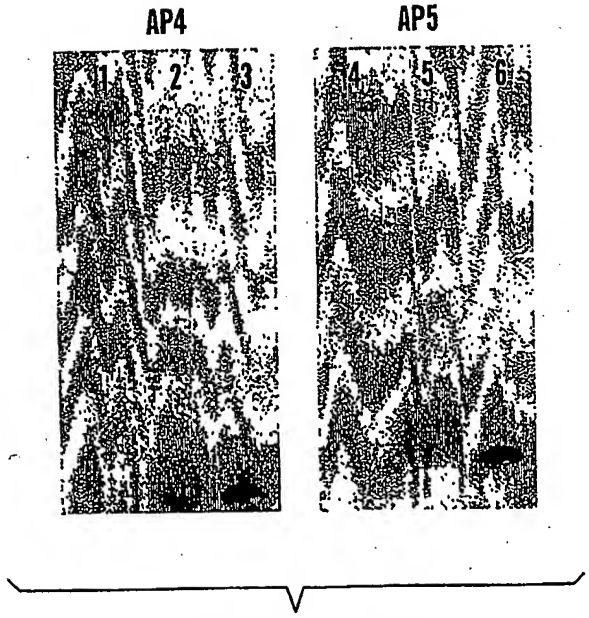


FIG. 8C

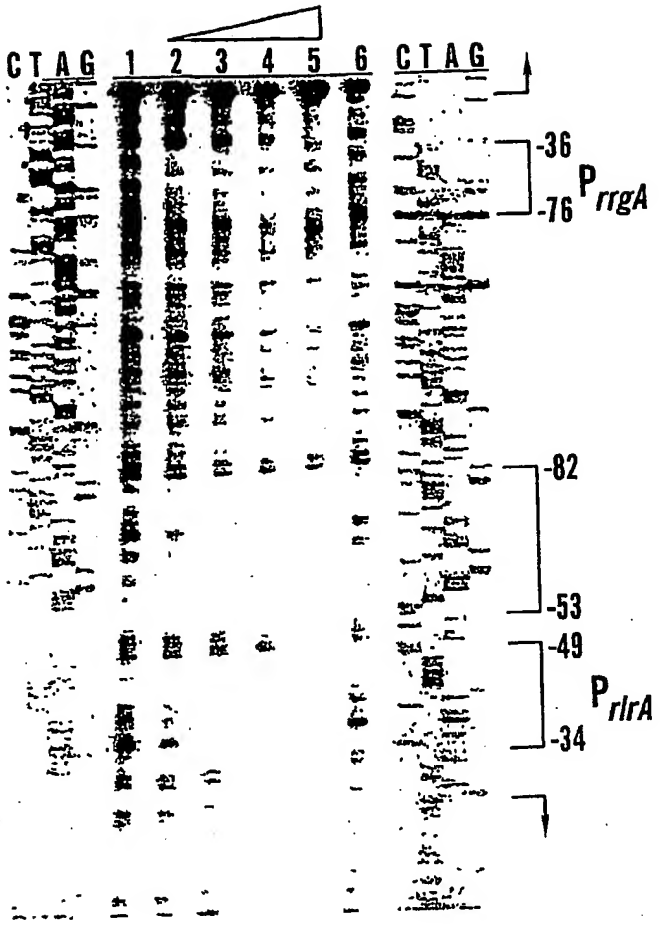

C T A G
1
2
3
4
5
6
C T A G
-36
-76
P rrgA
-82
-53
-49
-34
P rlrA

FIG. 9A

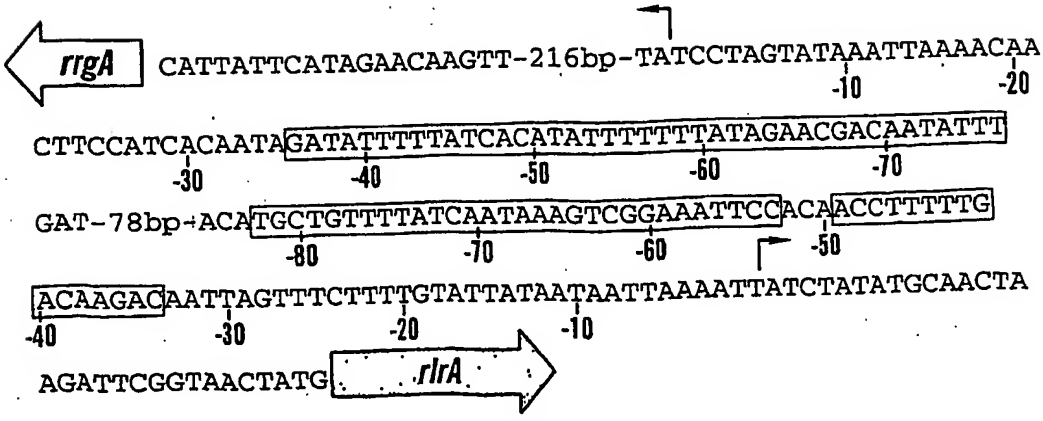

rrgA
CATTATTCATAGAACAAGTT-216bp-TATCCTAGTATAAATTAAAACAA
-10
-20
CTTCCATCACAATAGATATTTTTATCACATATTTTTTTATAGAACGACAATATTT
-30
-40
-50
-60
-70
GAT-78bp-ACATGCTGTTTTATCAATAAAGTCGGAAATTCCACAACCTTTTTG
-80
-70
-60
-50
ACAAGACAATTAGTTTCTTTTGTATTATAATAATTAAAATTATCTATATGCAACTA
-40
-30
-20
-10
AGATTCGGTAACTATG
rlrA

FIG. 9B